

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

OM nucleic - nucleic search, using sw model

Run on: September 2, 2004, 00:17:05 ; Search time 7415 Seconds
(without alignments)
11024.270 Million cell updates/sec

Title: US-09-423-575-1
Perfect score: 1886
Sequence: 1 attagtataaaattgtgc.....ttataaaaaaaaaaaaaa 1886

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl : *

1: gb_ba : *
2: gb_hgt : *
3: gb_in : *
4: gb_om : *
5: gb_ov : *
6: gb_ph : *
7: gb_pl : *
8: gb_pr : *
9: gb_ro : *
10: gb_ro : *
11: gb_ro : *
12: gb_sy : *
13: gb_un : *
14: gb_vt : *
15: em_ba : *
16: em_fun : *
17: em_hum : *
18: em_in : *
19: em_mu : *
20: em_om : *
21: em_or : *
22: em_ov : *
23: em_pat : *
24: em_ph : *
25: em_pi : *
26: em_ro : *
27: em_sts : *
28: em_un : *
29: em_vt : *
30: em_vt : *
31: em_hgt_inv : *
32: em_hgt_inv : *
33: em_hgt_inv : *
34: em_hgt_inv : *
35: em_hgt_inv : *
36: em_hgt_inv : *
37: em_hgt_inv : *
38: em_hgt_inv : *
39: em_hgt_inv : *
40: em_hgt_inv : *
41: em_hgt_inv : *

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB ID	Description
1	1886	100.0	1886	8	ATH1RNA
2	1882.8	99.8	1886	6	AX002854
3	1846	97.9	2277	8	AY072174
4	1476	78.3	9359	6	AX384394
5	1476	78.3	9359	6	AX473364
6	1453	77.0	1453	8	AY098513
7	1110.6	58.9	4878	8	ATH1GEN
8	1110.6	58.9	94482	8	ATF26P21
9	1110.6	58.9	198669	8	ATCHRIV78
10	702.8	37.3	10072	8	ATZ83320
11	255	13.5	1182	8	AF375967
12	132	7.0	2065	8	AF406703
13	128.6	6.8	2735	8	AF406703
14	127	6.7	2636	8	AF375966
15	126.4	6.7	2570	8	AK067939
16	126.4	6.7	2630	8	AK069994
17	126.2	6.7	1898	8	AF406698
18	125.6	6.7	2128	8	AF406702
19	124.2	6.6	2145	8	AK109849
20	124	6.6	1089	8	AK061393
21	124	6.6	1089	8	AK104713
22	123.6	6.6	2043	6	AX506642
23	123.6	6.6	2043	8	BT008422
24	123.6	6.6	2074	8	AY096548
25	123.6	6.6	2385	8	AF353094
26	123.6	6.6	2572	8	AY072175
27	123.6	6.6	2578	8	AY099854
28	122.6	6.5	1731	8	AF406700
29	121.8	6.5	1599	6	AX507459
30	121.4	6.4	2059	8	AY052366
31	119.8	6.4	1443	6	AX412914
32	119.8	6.4	1443	8	BT000609
33	119.8	6.4	2126	8	AY056796
34	117.8	6.2	1907	8	AF375964
35	115	6.1	1896	8	AK104872
36	115	6.1	2476	8	AK121356
37	115	6.1	3030	8	AK070465
38	114.8	6.1	1939	8	AF406701
39	114.4	6.1	1884	6	AX506210
40	114.4	6.1	1884	8	AT142019
41	114.4	6.1	2050	8	AF353092
42	114.4	6.1	2124	8	AY058179
43	114	6.0	2952	8	AK070543
44	113.4	6.0	2545	8	AF173816
45	113.4	6.0	3162	8	AY050459

ALIGNMENTS

RESULT 1

ATH1RNA

LOCUS ATH1RNA

DEFINITION A.thaliana Ath1 mRNA.

ACCESSION X80126

VERSION X80126.1

KEYWORDS H1 gene.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE 1

AUTHORS Quaadvliet, N., Dockx, J., Rook, F., Weisbeek, P. and Sneekens, S.

TITLE The homeobox gene ATH1 of Arabidopsis is derepressed in the

1886 bp mRNA linear PLN 02-MAR-1995

1321 GGATCTGTATGGAGAGAGGCAAGCAAGACTCAAGAAACCTCTATGTTCCACAGCAT 1380
 Db
 1381 TGCCTTCTTCCAGCAGCTGAACCAAGAAAGAACCATCATGATTTGGAGACCTCAACGAGTTTG 1440
 QY
 1381 TGCCTTCTTCCAGCAGCTGAACCAAGAAAGAACCATCATGATTTGGAGACCTCAACGAGTTTG 1440
 Db
 1441 CTGAGAAATCTGTTTCGGTTCTACGGAATTCGATGTTTCCAAAACCTCTTCCACCTTAC 1500
 QY
 1441 CTGAGAAATCTGTTTCGGTTCTACGGAATTCGATGTTTCCAAAACCTCTTCCACCTTAC 1500
 Db
 1501 CGGAAAGATTTCGGAGAAACATCTTCTAGCTATACCAAGTGGCTTGCACAAAGAGTCAAGTA 1560
 QY
 1501 CGGAAAGATTTCGGAGAAACATCTTCTAGCTATACCAAGTGGCTTGCACAAAGAGTCAAGTA 1560
 Db
 1561 TCAAACTGTTTATTAATGCGGGTTAGGCTATGGAAGCCGATGATGAGAGATGAT 1620
 QY
 1561 TCAAACTGTTTATTAATGCGGGTTAGGCTATGGAAGCCGATGATGAGAGATGAT 1620
 Db
 1621 GCGGAATCAACAGGAGGAGCTCAATACAGTCACTTCAACCAAGGACCAACTCTT 1680
 QY
 1621 GCGGAATCAACAGGAGGAGCTCAATACAGTCACTTCAACCAAGGACCAACTCTT 1680
 Db
 1681 CGAATGCCAAAATCTGTTATGATGACCAAGCAATGCATAATAGACAACAATTTGTT 1740
 QY
 1681 CGAATGCCAAAATCTGTTATGATGACCAAGCAATGCATAATAGACAACAATTTGTT 1740
 Db
 1741 TACCAACTTTGTAATTTAGGCAATGCTCTCTATGATGATGATGATGATGATGATGATGAT 1800
 QY
 1741 TACCAACTTTGTAATTTAGGCAATGCTCTCTATGATGATGATGATGATGATGATGATGAT 1800
 Db
 1801 TAGCACTATCATTCAGTATGTTATATGTTATATGTTATATGTTATATGTTATATGTTAT 1860
 QY
 1801 TAGCACTATCATTCAGTATGTTATATGTTATATGTTATATGTTATATGTTATATGTTAT 1860
 Db
 1861 ATTTTATTAATAAAAAAAAAAAAAA 1886
 QY
 1861 ATTTTATTAATAAAAAAAAAAAAAA 1886
 Db

RESULT 3
 AY072174 2277 bp mRNA linear PLN 18-SEP-2002
 LOCUS Arabidopsis thaliana putative homeobox gene ATH1 protein
 DEFINITION (At4g32980) mRNA, complete cds.
 ACCESSION AY072174
 VERSION AY072174.1 GI:18176167
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 2277)
 Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
 Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
 Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
 Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
 Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
 Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
 Davis,R.W., Ecker,J.R. and Theologis,A.
 Arabidopsis Full Length cDNA Clones
 Unpublished
 2 (bases 1 to 2277)
 Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M.,
 Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L.,
 Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,
 Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
 Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
 Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
 Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A.,
 Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
 Direct Submission
 Submitted (02-JAN-2002) Plant Gene Expression Center, 800 Buchanan

Street, Albany, CA 94710, USA
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA : RIKEN
 Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
 Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
 Hayashizaki,Y. and Shinozaki,K.
 The Salk, Stanford, PCEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Yamada,K., Banh,J.,
 Chan,M.M., Chang,C.H., Chang,E., Dale,J.M., Deng,J.M.,
 Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C.,
 Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,
 Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,
 Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P.,
 Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.
 Yamada,K. (SSP/PCEC) and Seki,M. (RIKEN GSC) contributed equally to
 this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PCEC)
 contributed equally to this work as PIs.
 Annotation is based on the January 2002 version of the Arabidopsis
 genome submitted to GenBank.
 FEATURES
 source
 1..2277
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="4"
 /clone="RAFL09-24-021 (R18486)"
 /ecotype="Columbia"
 /notes="This clone is in a modified pBluescript vector
 (FLC-1) as a BamHI/XhoI insert."
 gene
 1..2277
 /gene="At4g32980"
 5'UTR
 1..707
 /gene="At4g32980"
 misc_difference 2..3
 /gene="At4g32980"
 /notes="compared to genomic sequence"
 /replace="gta"
 misc_difference 373
 /gene="At4g32980"
 /notes="not present in genomic sequence"
 misc_difference 595
 /gene="At4g32980"
 /notes="not present in genomic sequence"
 /notes="not present in genomic sequence"
 708..2129
 /gene="At4g32980"
 /codon_start=1
 /evidence=experimental
 /product="putative homeobox gene ATH1 protein"
 /protein_id="AAU5996.1"
 /db_xref="GI:18176168"
 /translation="MDNNNNNTFSLDNVNTNQLPMDFFPSREDSTSFSTMLPWN
 TIRSDLOMGDFDIPNSMLTKYSSPSRIDQNRNVFMAPPHPPLPLDLHL
 HYDSSNNWGPANSEFOAFQSVGVPSPMSTGEEDFPELISKNRNEISLAS
 DVSECEISLCAATLASEQSCSKDISNNVTGQFSQILGSKYLHVSQVILSHF
 AAYSLDSRGTSAGAASTAFENITFEFDGNSNSEAGFGSTFQRALEAKTH
 LIDLLQVDRVSHCYDETHITISAFHAATELDPQLHTFALQVTFVLYKNRERICK
 KIISMGVLERGDKTQETSMFQHCQLQKLRKHQIWRPQRPGEKSVSVIRNMWF
 QNFHPYKDSKHLAIRSLRQSVNFINARVRLWPKMIEMEYAMNKRKLNS
 HIQNGPTLRMPKSVNMSQAMHK"
 2130..2277
 /gene="At4g32980"
 misc_difference 2261
 /gene="At4g32980"
 /notes="not present in genomic sequence"
 ORIGIN
 Query Match 97.9%; Score 1846; DB 8; Length 2277;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 1863; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 ATTAGTATAAATGTTGCTATTTGTTGATCTAGTCTGAACTTTTGTAGTGAGCAG 60
Db ATTTAGTATAAATGTTGCTATTTGTTGATCTAGTCTGAACTTTTGTAGTGAGCAG 463
QY 61 ATGATGAAGATTATGAATTTCTCATGAATTTATTTGTAAGAAAAGAAATAGAGAGCT 120
Db ATGATGAAGATTATGAATTTCTCATGAATTTATTTGTAAGAAAAGAAATAGAGAGCT 523
QY 121 GCGGAATGAAGTACACTGTTCTTTTCCAGGAGAAAAGATAAATAAGCAATTATCTCTT 180
Db GCGGAATGAAGTACACTGTTCTTTTCCAGGAGAAAAGATAAATAAGCAATTATCTCTT 583
QY 181 CTTTCAG-TTTTAAACACACATTTTGGAAATTTTGAATTTTGAATTTTCTTTGGAGAGTTG 239
Db CTTTCAGTTTAAACACACATTTTGGAAATTTTGAATTTTGAATTTTCTTTGGAGAGTTG 643
QY 240 TGTGTTCTGAAATCTTCCCAAGGTTCTATCAAGAGAAAGAGATATAAGTTTTCATAGAAA 299
Db TGTGTTCTGAAATCTTCCCAAGGTTCTATCAAGAGAAAGAGATATAAGTTTTCATAGAAA 703
QY 300 CCAATATGGACAAACAAACAAACACTTTTATTTCTGTGATAATGTATGACTATA 359
Db CCAATATGGACAAACAAACAAACACTTTTATTTCTGTGATAATGTATGACTATA 763
QY 360 CCAAAATCTCTCTCATGATTTTATACCTTCAAGAGAGATTCAACTTCATCTCAAC 419
Db CCAAAATCTCTCTCATGATTTTATACCTTCAAGAGAGATTCAACTTCATCTCAAC 823
QY 420 AATGCTTCCATGAAATACCATCAGATCAGATCCTCTACAAATGGGTGGCTTTGATATTT 479
Db AATGCTTCCATGAAATACCATCAGATCAGATCCTCTACAAATGGGTGGCTTTGATATTT 883
QY 480 CAATCTATGCTGACTAAACAAATACCTTATCACTTCTCCACGGTCTATCGATGTTCAAGA 539
Db CAATCTATGCTGACTAAACAAATACCTTATCACTTCTCCACGGTCTATCGATGTTCAAGA 943
QY 540 TAAACCAATGTTGAGTTTATGGCTCTCTCTCATCTCTCTCCACATTCATCTCTTGA 599
Db TAAACCAATGTTGAGTTTATGGCTCTCTCTCATCTCTCTCCACATTCATCTCTTGA 1003
QY 600 TCATTTAAGACATATGATGATTCCTCAACAAACATGTTGGGTTTTGAAGCAAAATAGTGA 659
Db TCATTTAAGACATATGATGATTCCTCAACAAACATGTTGGGTTTTGAAGCAAAATAGTGA 1063
QY 660 GTTTCAGGCAATTTTCAGGTGATGTTGGTCCAAAGTGAACCAATGATGCTACATTCGGTGA 719
Db GTTTCAGGCAATTTTCAGGTGATGTTGGTCCAAAGTGAACCAATGATGCTACATTCGGTGA 1123
QY 720 AGAAGATTTCCGTTTCTAATTTTGAATAAAGAAAACAAATGAGCTTTTCAATGAGTCTTGC 779
Db AGAAGATTTCCGTTTCTAATTTTGAATAAAGAAAACAAATGAGCTTTTCAATGAGTCTTGC 1183
QY 780 ATCAGATGTTTCTGATGATGCTCGGAGATAGTCTTTGTCGAGCTACAGATTAGCCTC 839
Db ATCAGATGTTTCTGATGATGCTCGGAGATAGTCTTTGTCGAGCTACAGATTAGCCTC 1243
QY 840 AGAGCAAGCTTCTTGACAGCAAGAACATTTCTAATAACGTTGTTTACTCAAGGTTTCTC 899
Db AGAGCAAGCTTCTTGACAGCAAGAACATTTCTAATAACGTTGTTTACTCAAGGTTTCTC 1303
QY 900 TCAACTTATATTGGCTCAAAATACCTTCACTCTGTTTCAAGAAAATACATCTCATTTGCG 959
Db TCAACTTATATTGGCTCAAAATACCTTCACTCTGTTTCAAGAAAATACATCTCATTTGCG 1363
QY 960 CGCATACTCGCTCGATTTATCTCTCGAGGAACCGAGTCAGAGCTGCTAGTTCAGCCTT 1019
Db CGCATACTCGCTCGATTTATCTCTCGAGGAACCGAGTCAGAGCTGCTAGTTCAGCCTT 1423
QY 1020 TACTTCAGCTTTTGAATAATACTGAGTTTCTTGATGTTGATTTCTTAATAACTCGAGGC 1079
Db TACTTCAGCTTTTGAATAATACTGAGTTTCTTGATGTTGATTTCTTAATAACTCGAGGC 1483
QY 1080 GGGTTTCGGATCTACATTTCAAGAGGAGACATTTAGAGCAAGAAAACCCATCTCTTGA 1139

Db 1484 GGGTTTCGGATCTACATTTCAAGAGGAGACATTAGAGCAAGAAAACCCATCTCTTGA 1543
QY 1140 TCTTCTTCAATGTTGGATGATCGATATAGTCAATTCATTCGGTAGATGAGATTCATACGGTTAT 1199
Db 1544 TCTTCTTCAATGTTGGATGATCGATATAGTCAATTCATTCGGTAGATGAGATTCATACGGTTAT 1603
QY 1200 ATCAGCGTTCCATGCTGCAACCGAGTTAGATTCACAGTTACACACCCGGTTTGGCCCTCCA 1259
Db 1604 ATCAGCGTTCCATGCTGCAACCGAGTTAGATTCACAGTTACACACCCGGTTTGGCCCTCCA 1663
QY 1260 AACCGTTTCTTCTTATACAAAGAACCTTGAGAGAGAGATCTGCAAGAGAGATAATCTCTAT 1319
Db 1664 AACCGTTTCTTCTTATACAAAGAACCTTGAGAGAGAGATCTGCAAGAGAGATAATCTCTAT 1723
QY 1320 GGGATCTGTATTGGAGAGAGGCAAGACACTCAAGAAACCTCTATGTTTCCACACAGCA 1379
Db 1724 GGGATCTGTATTGGAGAGAGGCAAGACACTCAAGAAACCTCTATGTTTCCACACAGCA 1793
QY 1380 TTGCTCTTTCAGCAGCTGAAACGAAAGAACCAATCAGATTTGGAGAGCTCAACAGAGTTT 1439
Db 1784 TTGCTCTTTCAGCAGCTGAAACGAAAGAACCAATCAGATTTGGAGAGCTCAACAGAGTTT 1843
QY 1440 GCCTGAGAAATCTGTTTCGGTTCTACGGAATGGAATGTTCCAAACTTCTTCCACCTTA 1499
Db 1844 GCCTGAGAAATCTGTTTCGGTTCTACGGAATGGAATGTTCCAAACTTCTTCCACCTTA 1903
QY 1500 CCCGAAAGATTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGAAGAACTCAGGT 1559
Db 1904 CCCGAAAGATTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGAAGAACTCAGGT 1963
QY 1560 ATCAAACTGTTTATAATGCGGGTTAGCTATGGAAGCCGATGATAGAGAGATGTA 1619
Db 1964 ATCAAACTGTTTATAATGCGGGTTAGCTATGGAAGCCGATGATAGAGAGATGTA 2023
QY 1620 TCGGAAATGAACAGAGAGAGCTCAATAACAGTCAATTCACATTCACCCCAACGACCACTCT 1679
Db 2024 TCGGAAATGAACAGAGAGAGCTCAATAACAGTCAATTCACATTCACCCCAACGACCACTCT 2083
QY 1680 TCGAATGCCAAATCTGTTATGATGAGCCAAAGAAATGCAATAAAGACAAATTTGTTG 1739
Db 2084 TCGAATGCCAAATCTGTTATGATGAGCCAAAGAAATGCAATAAAGACAAATTTGTTG 2143
QY 1740 TTACCACTTTGTGATTAATTAGGCAATTTGCTACTCTATGATTGCCCAAAACCTTAAACCAT 1799
Db 2144 TTACCACTTTGTGATTAATTAGGCAATTTGCTACTCTATGATTGCCCAAAACCTTAAACCAT 2203
QY 1800 GTACGATATCATTAAGTATGTTAATTTGATATATACAACTCTTTTATCTTTGACTATTT 1859
Db 2204 GTACGATATCATTAAGTATGTTAATTTGATATATACAACTCTTTTATCTTTGACTAGAA 2263
QY 1860 CATTTTATTAATA 1873
Db 2264 AAAAAAAAAAAAAA 2277

RESULT 4
AX384394LOCUS AX384394 9359 bp DNA linear PAT 19-MAR-2002
DEFINITION Sequence 3 from Patent WO0214524.ACCESSION AX384394
VERSION AX384394.1 GI:19577678KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.REFERENCE
AUTHORS van der Valk, P., van Dun, C.M., Smeekens, S.C. and Proveniers, M.C.TITLE Inhibition of generative propagation in genetically modified
herbicide resistant grassesJOURNAL Patent: WO 0214524-A 3 21-FEB-2002;
Advanta Seeds B.V. (NL)FEATURES
Location/Qualifiers

Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
Arabidopsis Open Reading Frame (ORF) Clones

JOURNAL REFERENCE AUTHORS

Unpublished
2 (bases 1 to 1453)

Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, J.M., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
Direct Submission

TITLE JOURNAL

Submitted (16-APR-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

COMMENT

The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA : RIKEN Arabidopsis Full-length cDNA): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and sequenced the PUNI (ORF) clones using the RAPL cDNAs: Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, J.M., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis genome submitted to GenBank.

FEATURES source

Location/Qualifiers
1..1453
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="4"
/clone="U18486"
/ecotype="Columbia"
/note="this clone is in PUNI 51."
1..1453
/gene="At4g32980"
1..1422
/gene="At4g32980"
/codon_start=1
/evidence="experimental"
/product="putative homeobox gene ATH1 protein"
/protein_id="AA020163.1"
/db_xref="GI:20465955"
/translation="MDNNNNNTTSSLDNVNNTQNPMLDIPREDSTSFSTMLPWN
TIRSDPLQMGDFDINSMNTKYLSSSPSIDQDNVEMFAPPPHPLPLDHLR
HYDSSNNMGFEANSEFCFVGVVPEPMMSTFGEDRPFLI SNKRNELSLSLAS
DVSDECSISLCATRLASFOASGSDISNNVTOGFSOLI FGSKYLHSVQETLSHF
RAYSLDYSRSTESGAASAFSTSPENITFELDGSNNSEAGFSGSTFORALEAKYTH
LDDLQVDDIRSHCVBIHTVIFAFHAATLDFQLHTRFALQTVSLYKLNLERICK
KIISGVLSEKQKTOBTSMFQHLQLQKLRKHQIWRPQGLPKSVSLRNWMP
QNFLPYPKDSEKHLAIRSGITRSQVSNWFINARVLRWMEIEMVAEMVNRKLNN
HIQNGPTRLRVPKSVWMSQAMHK"
1423..1453
/gene="At4g32980"

gene

CDS

3'UTR

ORIGIN

Query Match 77.0%; Score 1453; DB 8; Length 1453;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

304 ATGGACAACAAACAAACAAACAAACAACTTTTAGTTCTCTGGTAATGTCATGACTAACCAA 363
1 ATGGACAACAAACAAACAAACAAACAACTTTTAGTTCTCTGGTAATGTCATGACTAACCAA 60
364 AATCCTCTTCTCATGATTTTATACCTTCAAGAGAAGATTCACTTCATCTTCAACAATG 423
61 AATCCTCTTCTCATGATTTTATACCTTCAAGAGAAGATTCACTTCATCTTCAACAATG 120
424 CTTCCATGGAATACCAATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAGAT 483
121 CTTCCATGGAATACCAATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAGAT 180
484 TCTATGCTGATCAACAAATACCTTATCATCTTCTCAAGGTCTATCGATGTTCAAGATAAC 543
181 TCTATGCTGATCAACAAATACCTTATCATCTTCTCAAGGTCTATCGATGTTCAAGATAAC 240
544 CGCAATGTTGAGTTTCAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 603
241 CGCAATGTTGAGTTTCAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
604 TTAAGACATATGATGATTCCTCAAAACAACATGTGGGTTTTGAAGCAAAATAGTGAGTTT 663
301 TTAAGACATATGATGATTCCTCAAAACAACATGTGGGTTTTGAAGCAAAATAGTGAGTTT 360
664 CAGGCATTTTCAGGTGAGTTCGTCCAGTGAACCAATGATGCTTACATTCGGTGAAGAA 723
361 CAGGCATTTTCAGGTGAGTTCGTCCAGTGAACCAATGATGCTTACATTCGGTGAAGAA 420
724 GATTTCCCGTTTCTAAATTCGAATTAAGAAACAATAGAGCTTTCAATGAGTCTTCATCA 783
421 GATTTCCCGTTTCTAAATTCGAATTAAGAAACAATAGAGCTTTCAATGAGTCTTCATCA 480
784 GATTTTCTGATGATGCTCGAGATAAGTCTTTGTGAGCTACAGATTAGCCTCAGAG 843
481 GATTTTCTGATGATGCTCGAGATAAGTCTTTGTGAGCTACAGATTAGCCTCAGAG 540
844 CAAGCTTCTTGACGACGACAAAGACATTTCTAATAAGTTGTTTACTCAAGGTTTCTCTCAA 903
541 CAAGCTTCTTGACGACGACAAAGACATTTCTAATAAGTTGTTTACTCAAGGTTTCTCTCAA 600
904 CTTATATTTGGCTCAAAATACCTTCACTCTGTTTCAAGAAATACCTATCTCATTTCCGCGCA 963
601 CTTATATTTGGCTCAAAATACCTTCACTCTGTTTCAAGAAATACCTATCTCATTTCCGCGCA 660
964 TACTCGCTCGATTTATTCATCTCGAGAACCGAGTCAAGAGCTGCTAGTTACAGCCTTTACT 1023
661 TACTCGCTCGATTTATTCATCTCGAGAACCGAGTCAAGAGCTGCTAGTTACAGCCTTTACT 720
1024 TCACGTTTTCAGATAATAAATCAGTTTCTTGTGATGTTTCTAATAACTCGAGGCGGGT 1083
721 TCACGTTTTCAGATAATAAATCAGTTTCTTGTGATGTTTCTAATAACTCGAGGCGGGT 780
1084 TTCGGATCTACATTTCAAGGAGAGGATTTAGAGCAAGAAACCACTCTCTTTGATCTT 1143
781 TTCGGATCTACATTTCAAGGAGAGGATTTAGAGCAAGAAACCACTCTCTTTGATCTT 840
1144 CTTCAAAATGGTGGATCGATATAGTCAATTCGCTAGATGAGATTCATACCGTTTATATCA 1203
841 CTTCAAAATGGTGGATCGATATAGTCAATTCGCTAGATGAGATTCATACCGTTTATATCA 900
1204 GCGTTCCATGCTGCACCGAGTTAGTCCACAGTTTACACACCGGTTTGCCTCCCAACC 1263
901 GCGTTCCATGCTGCACCGAGTTAGTCCACAGTTTACACACCGGTTTGCCTCCCAACC 960
1264 GTTTCCTTCTTATACAAGAACCTTGAGAGAGAATCTGCAAGAAAGATATCTTCTATGGA 1323
961 GTTTCCTTCTTATACAAGAACCTTGAGAGAGAATCTGCAAGAAAGATATCTTCTATGGA 1020
1324 TCTGTATTCGAGAGAGGCAAGAAAGAACTTCTATGTTCCAGCAGATTGC 1383
1021 TCTGTATTCGAGAGAGGCAAGAAAGAACTTCTATGTTCCAGCAGATTGC 1080
1384 CTTCTTTCAGAGCTGAAACGAAAGAAACCATCAGATTTCGAGACCTCAACAGGTTTGCT 1443

763 CTTTCATTGAGTCTTGCATCAGATGTTTCTGATGAATGCTCGAGATAAGTCTTTTGCA 822
Db 2940 CTTTCATTGAGTCTTGCATCAGATGTTTCTGATGAATGCTCGAGATAAGTCTTTTGCA 2999
Qy 823 GCTACAAGATTAGCTCAGAGCAAGCTTCTTCAGCAGCAAGACATTTCTAATAAGTT 882
Db 3000 GCTACAAGATTAGCTCAGAGCAAGCTTCTTCAGCAGCAAGACATTTCTAATAAGTT 3059
Qy 883 GTTACTCAAGTTTCTCTCAACTTATTTGGCTCAAAATACCTTCACTCTGTTCAAGAA 942
Db 3060 GTTACTCAAGTTTCTCTCAACTTATTTGGCTCAAAATACCTTCACTCTGTTCAAGAA 3119
Qy 943 ATACTATCTCAATTTCCGCGCATACTCGTCGATTAATTCATCTCAGAGAACCGAGTCAAG 1002
Db 3120 ATACTATCTCAATTTCCGCGCATACTCGTCGATTAATTCATCTCAGAGAACCGAGTCAAG 3179
Qy 1003 GCTGCTAGTTCAGCCTTTTACTTTCAGTTTTCAGATATACTGAGTTTCTTGTGTTGAT 1062
Db 3180 GCTGCTAGTTCAGCCTTTTACTTTCAGTTTTCAGATATACTGAGTTTCTTGTGTTGAT 3239
Qy 1063 TCTAATACTCGAGCGGGTTTCGGATCTACATTTCAAGGAGAGCATTAAGAGCAAG 1122
Db 3240 TCTAATACTCGAGCGGGTTTCGGATCTACATTTCAAGGAGAGCATTAAGAGCAAG 3299
Qy 1123 AAAACCCATCTCTCGATCTTCTCAAT 1151
Db 3300 AAAACCCATCTCTCGATCTTCTCAAT 3359
Qy 1152 1152 1152
Db 3360 ATAGTTAAAAAATGTTTGATATATATAGTAAGTTTATGAGTTGATATTTCTCTTC 3419
Qy 1152 -GGTGGATCGATATAGTCAATTCGGTAGATGAGATCAATACGGTTATATACAGGTTCC 1210
Db 3420 AGGTGGATCGATATAGTCAATTCGGTAGATGAGATCAATACGGTTATATACAGGTTCC 3479
Qy 1211 ATGCTGCAACCGAGTTAGATCAACAGTTACACACCGGTTTGCCTCCAAACCGTTTCC 1270
Db 3480 ATGCTGCAACCGAGTTAGATCAACAGTTACACACCGGTTTGCCTCCAAACCGTTTCC 3539
Qy 1271 TCTTATACAAGAACTGAGAGAGAAATCTGCAAGAGATATCTCTATGGGATCTGTAT 1330
Db 3540 TCTTATACAAGAACTGAGAGAGAAATCTGCAAGAGATATCTCTATGGGATCTGTAT 3599
Qy 1331 TGGAGAGGCAAGAGACAGCTCAAGAACTCTATGTTCCACAGCATTCCTTCTTC 1390
Db 3600 TGGAGAGGCAAGAGACAGCTCAAGAACTCTATGTTCCACAGCATTCCTTCTTC 3659
Qy 1391 AGCAGCTGAACGAAGAACCATCAGATTGAGACCTCAACGAGTTTGCCTGAGAAAT 1450
Db 3660 AGCAGCTGAACGAAGAACCATCAGATTGAGACCTCAACGAGTTTGCCTGAGAAAT 3719
Qy 1451 CTTGTTTGGTTCTACGGAATGGATGTTCCAAACCTTCTTCAACCTTACC 1501
Db 3720 CTTGTTTGGTTCTACGGAATGGATGTTCCAAACCTTCTTCAACCTTACC 3770

RESULT 8
LOCUS AT26P21/c
DEFINITION Arabidopsis thaliana DNA linear PLN 16-MAR-2000 project.
ACCESSION AL031804
VERSION AL031804.1 GI:3688169
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 94482)
AUTHORS Bevan, M., Terry, N., Ardiles, W., Buyssehaert, C., Dasseville, R., De

Clerck, R., De Keyser, A., Neyt, P., Rouze, P., Van Den Daele, H., Villard, R., Gielens, J., Van Montagu, M., Hohn, J., Mewes, H. W., Lemcke, K. and Mayer, K. F. X.
Unpublished
2 (bases 66824 to 67623; 872086 to 72704; 87789 to 88522; 91317 to 92029)
Bevan, M., Volckaert, G., Grymopoulos, B., Voet, M., Robben, J., Hohn, J., Mewes, H. W., Lemcke, K. and Mayer, K. F. X.
Unpublished
3 (bases 1 to 94482)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mpibiochem.mpg.de, mayer@mpibiochem.mpg.de
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk
Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.
Location/Qualifiers
1. 94482
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
6659..7676
/gene="AT4G32890"
join(6659..6877,6969..7676)
/gene="AT4G32890"
/note="similarity to GATA transcription factor 3, Arabidopsis thaliana, gb:Y13650
contains EST gb:AA041075, AI994034.1, AI994545.1, F14085"
/codon_start=1
/product="putative protein"
/protein_id="CAA21198.1"
/db_xref="GI:3688170"
/db_xref="GOA:O82632"
/db_xref="SPTREMBL:O82632"
/translation="MEKLAPELFLVAGNPDSPVDDLLDFNSDGDGVDGLNTLPDSS TSTGLTDSNSSLFTDGTGFSGLYIPNDIALEWLSNFKVRSFAGEQDQXHLFL SGLKPPQTTGTLTHLIRPELPHQFIDIBSNVAVPAKRSRASAATWASRL SLADENPTNKKKQRRVKEQDFAGMDVDCGGGRCCLCHCATKTKTQWRITGPGPK TILCNAGVRYKSGRLVPEYRSPASPTFVYARHSNHRKVMELRRQKEMRDEHLLSQLR CENLLMDIRNSGDFLMHNNTNHNVPDFRHLI"
6659..6877
/gene="AT4G32890"
/number=1
6878..6968
/gene="AT4G32890"
/number=1
6969..7676
/gene="AT4G32890"
/number=2
10392..12039
/gene="AT4G32900"
complement(join(10392..10683,10773..10809,10882..10942, 11225..11425,11998..12039))
/gene="AT4G32900"
complement(join(10392..10683,10773..10809,10882..10942, 11225..11425,11998..12039))
/gene="AT4G32900"
/note="Contains Sugar transport proteins signatures AA31-56
contains EST gb:AI995200.1, N37475, AA712862"
/codon_start=1
/product="putative protein"
/protein_id="CAA21199.1"
/db_xref="GI:3688171"
/db_xref="GOA:O82633"
/db_xref="SPTREMBL:O82633"

exon	/number=1 9225..9303 /gene="AT4g32720"	WMSNDGESSRFGVDAERSESCQASTVFSQSTNDLQDEVQORNEEYMPFHSQTEQ
intron	/number=2 9304..9402 /gene="AT4g32720"	QISNAASHAEPYYPFKDKIVVPEISCEETESCKFQNLNCSELRTTATEDQLPGV
exon	/number=2 9403..9552 /gene="AT4g32720"	SNDAKQDRGLELLTHNMDGKGNALQQDFQSSVRLSDQFLNSDTPPEAQLITDE
intron	/number=3 9553..9829 /gene="AT4g32720"	ECCRVLFPDNNKDSSTSSGEQGRNVDPNQKSLCSQAAETHAHTGKVPALPWHPS
exon	/number=3 9830..9972 /gene="AT4g32720"	SSGLAGHNCVPLLLDSDLPRNDSNAPIQGCRLFGATELECKTDTNDGFDITYG
intron	/number=4 9973..10017 /gene="AT4g32720"	HVTSHGNDNGGPFQOGLSYIPKDSLKLVPILNSFSSPSRVNKIYFPIDDKPAEKDKG
exon	/number=4 10018..10211 /gene="AT4g32720"	ALCYEPRFSAIPIFFSCDLVFSNLDLQEVSPFGIRQLMISSMNCATPLRLWDSPC
intron	/number=5 10212..10324 /gene="AT4g32720"	HDRSPVNLMDTAKSFGAPSLIKKRRHLLSPVLDKRDKKLKRATTSGLANDFRSL
exon	/number=5 10325..10447 /gene="AT4g32720"	DVNLDEGDDCMTSPESPEDKNICASPSIARDNRNCAARLYQEWIPIDEEPKEITLE
intron	/number=6 10448..10536 /gene="AT4g32720"	SGGVTSNQENGCNDGASAKNVSPSLHLHIWYQL"
exon	/number=6 10537..10831 /gene="AT4g32720"	/gene="AT4g32730"
intron	/number=7 10832..10936 /gene="AT4g32720"	/number=1
exon	/number=7 10937..11084 /gene="AT4g32720"	12868..13003
intron	/number=8 11085..11144 /gene="AT4g32720"	/gene="AT4g32730"
exon	/number=8 11145..11262 /gene="AT4g32720"	/number=1
intron	/number=9 11263..11306 /gene="AT4g32720"	13004..13058
exon	/number=9 11307..11457 /gene="AT4g32720"	/gene="AT4g32730"
gene	/number=10 12797..16010 /gene="AT4g32730"	/number=2
CDS	Join(12797..12867,13004..13058,13283..13340,13507..13601, 13702..13805,13902..14030,14192..16010) /gene="AT4g32730"	13059..13282
	/notes="similarity to transforming protein A-myb, Xenopus laevis, PIRI:S36095 Contains Myb DNA-binding domain repeat signatures A29-37;Myb DNA-binding domain repeat signatures A481-89;Myb DNA-binding domain repeat signatures AA49-72 contains EST gb:AW004163.1" /codon_start=1 /product="putative myb-protein" /protein_id="CA87990.1" /db_xref="GI:7270220" /db_xref="GGA:Q9S7G7" /db_xref="SPTREMBL:Q9S7G7" /translation="MKREMKAPTPLESLQDGLKGKQSTSGPARSTKGQWTPPEDE VLCKAVERFQGNKKIAECFDRIDVQCLHRWQVLPDELVKGPWSKEEDNTIIDLV EYKVPKWTISQHLPGRIKOCRRWHLNPNKNAWTOBELTILIRAHQIYGNK WAEILMKFLGRSDNSIKNHNSSVKKLDSYVASGLLDQCQSSPLIALQNKSIASSS	13283..13340

QY	643	TTTGAAGCAAAATAGTGGATTTTCCAGGATTTTCCAGGTGATGTTGGTCCAAGTGAACCAATG	702
Db	138290	TTTGAAGCAAAATAGTGGATTTTCCAGGATTTTCCAGGTGATGTTGGTCCAAGTGAACCAATG	138231
QY	703	ATGCTACATTTGGTGAAGAAATTTCCCGTTTCTAATTTGGAATTAAGAAACCAATGAG	762
Db	138230	ATGCTACATTTGGTGAAGAAATTTCCCGTTTCTAATTTGGAATTAAGAAACCAATGAG	138171
QY	763	CTTTCAATTCAGTCTTCATCAGATGTTTCTGATGAATGCTCCGAGATAAGTCTTTGTGCA	822
Db	138170	CTTTCAATTCAGTCTTCATCAGATGTTTCTGATGAATGCTCCGAGATAAGTCTTTGTGCA	138111
QY	823	GCTACAAGATTTAGCTTCAGAGCAAGTTTCTTCAGAGCAAGAAAGATTTCTAATAACGTT	882
Db	138110	GCTACAAGATTTAGCTTCAGAGCAAGTTTCTTCAGAGCAAGAAAGATTTCTAATAACGTT	138051
QY	883	GTCTACTCAAGGTTTCTCTCAACTTATTTGCTCAAAATACCTTCACTCTGTTCAAGAA	942
Db	138050	GTCTACTCAAGGTTTCTCTCAACTTATTTGCTCAAAATACCTTCACTCTGTTCAAGAA	137991
QY	943	ATACTATCTCATTTCCGCCATPACTCGCTCGATTATTCATCTCGAGGAACCGAGTCAGGA	1002
Db	137990	ATACTATCTCATTTCCGCCATPACTCGCTCGATTATTCATCTCGAGGAACCGAGTCAGGA	137931
QY	1003	GGTGTAGTTAGCTTTTACCTTACGTTTGGATATTAATCTGATGTTCTTTGATGGTAT	1062
Db	137930	GGTGTAGTTAGCTTTTACCTTACGTTTGGATATTAATCTGATGTTCTTTGATGGTAT	137871
QY	1063	TCTATACTCGAGCGGGTTTCCGATCTCATTTCAAAGAGAGCATTAAGAACCAAG	1122
Db	137870	TCTATACTCGAGCGGGTTTCCGATCTCATTTCAAAGAGAGCATTAAGAACCAAG	137811
QY	1123	AAACCCATCTCTTGATCTTCTTCAAT	1151
Db	137810	AAACCCATCTCTTGATCTTCTTCAAT	137751
QY	1152	-----	1151
Db	137750	ATAAGTTAAAAAATGGTTTGATATATATGATGAAGTTTATGAGTGAATATTTCTCTTC	137691
QY	1152	-GGTGAATGATCGATATAGTCAATTCGGTAGATGAGATTCATACGGTTATATCAGGTTCC	1210
Db	137690	AGGTGATGATCGATATAGTCAATTCGGTAGATGAGATTCATACGGTTATATCAGGTTCC	137631
QY	1211	ATGCTGCACCCAGTTAGATCCACAGTTACACACCGGTTTCCCTCCAAACCGTTTCCCT	1270
Db	137630	ATGCTGCACCCAGTTAGATCCACAGTTACACACCGGTTTCCCTCCAAACCGTTTCCCT	137571
QY	1271	TCTTATACAAGAACCTTGAGAGAGAGATCTGCAAGAGATAATCTCTATGGGATCTGTAT	1330
Db	137570	TCTTATACAAGAACCTTGAGAGAGAGATCTGCAAGAGATAATCTCTATGGGATCTGTAT	137511
QY	1331	TGAGAGAGGCAAGACACAGTCAAGAACCTCTATGTTCCACAGCATTCCTCTTCTTC	1390
Db	137510	TGAGAGAGGCAAGACACAGTCAAGAACCTCTATGTTCCACAGCATTCCTCTTCTTCTTC	137451
QY	1391	AGCAGTCAAGCAAGAAACCACTAGATTTGGAGACCTCAACGAGGTTTGGCTGAGAAAT	1450
Db	137450	AGCAGTCAAGCAAGAAACCACTAGATTTGGAGACCTCAACGAGGTTTGGCTGAGAAAT	137391
QY	1451	CTGTTTCGGTTCTACGGAATGGATGTTCCAAACCTTCCCTCACCCTTACC	1501
Db	137390	CTGTTTCGGTTCTACGGAATGGATGTTCCAAACCTTCCCTCACCCTTACC	137340

RESULT 10
 AT283320
 LOCUS AT283320
 DEFINITION A.thaliana partial ATH1 gene.
 ACCESSION Z83320
 VERSION Z83320.1 GI:1865680
 KEYWORDS ATH1 gene; MAR.

SOURCE ORGANISM	Arabidopsis thaliana (thale cress)
REFERENCE	Arabidopsis thaliana
AUTHORS	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
JOURNAL	1 (bases 1 to 10072)
MSDLINE	van Drunen,C.M., Oosterling,R.W., Keultjes,G.M., Weisbeek,P.J., van Driel,R. and Sneekens,S.C.
PUBMED	Analysis of the chromatin domain organisation around the plastocyanin gene reveals an MAR-specific sequence element in Arabidopsis thaliana
REFERENCE	Nucleic Acids Res. 25 (19), 3904-3911 (1997)
AUTHORS	98026209
TITLE	2 (bases 1 to 10072)
JOURNAL	Oosterling,R.W.
FEATURES	Direct Submission
source	Submitted (11-DEC-1996) Oosterling R.W., Utrecht University, Molecular cell Biology, Padualaan 8, Utrecht, The Netherlands, 3584 CH
Location/Qualifiers	
1..10072	/organism="Arabidopsis thaliana"
	/mol_type="genomic DNA"
	/variety="Columbia-0"
	/db_xref="taxon:3702"
	/chromosome="4"
	/clone="PH1Pr"
	/clone_lib="Lambda ZAP genomic library"
1..1431	/notes="ATH1 locus MAR-1"
	/evidence=experimental
471..489	/notes="ATH1 locus MAR-2"
	/evidence=experimental
2046..3762	/notes="ATH1 locus MAR-3"
	/evidence=experimental
7905..9103	/notes="ATH1 locus MAR-4"
	/evidence=experimental
9388..10072	/gene="ATH1"
	/gene="ATH1"
	/note="Related sequences: ATH1 genomic clone, acc. no. X80126"
	/codon_start=1
	/protein_id="CAB05910.1"
	/db_xref="GI:1865681"
	/db_xref="GOA:P48731"
	/db_xref="SWISS-PROT:P48731"
	/translation="MDNNNNNTFSSLDNVTNQNPMLMDIPSRDSTSFSTMLPWN TIRSPLOWGGDIENSMLTNKYLSSPRSIDVDNRNVEFMAPPHPPLDHLR HYDSSNNWGFANSEFQAGSVGVGPSEPMSTFGEEDFFPLISKNENNELSLAS DVSDESLISLCATRLASEQASCSKDISNNVTQGFSLIFGSKYLHSVQELIUSHP AAYSIDYSRR"
Query Match	37.3%; Score 702.8; DB 8; Length 10072;
Best Local Similarity	99.7%; Pred. No. 4.8e-149;
Matches 704; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
QY	283 ATAAAGTTTCATAGAAACCAATGAGCAACAAACAAACAAACACTTTAGTTCTCTG 342
Db	9367 ATGCAGTTTCATAGAAACCAATGAGCAACAAACAAACAAACACTTTAGTTCTCTG 9426

QY 343 GATAATGTCATGACTAACCAAAATCTCTCTCTCATGGATTTTATACCTTCAAGAGAAGAT 402
Db 9427 GATAATGTCATGACTAACCAAAATCTCTCTCTCATGGATTTTATACCTTCAAGAGAAGAT 9486
QY 403 TCAACTTCACTTCAACAATGCTTCCATGGAATACATCAGATCAGATCCTCTCAAAATG 462
Db 9487 TCAACTTCACTTCAACAATGCTTCCATGGAATACATCAGATCAGATCCTCTCAAAATG 9546
QY 463 GTGGCTTTGATATTTCATTTCAATTTCTATGCTGACTAAACAAATCTTATCATCTTCTCCACGG 522
Db 9547 GTGGCTTTGATATTTCATTTCAATTTCTATGCTGACTAAACAAATCTTATCATCTTCTCCACGG 9606
QY 523 TCTATCGATGTTCAAGATAACCGCAATGTTAGTTTCATGGCTCCTCCCTCATCTCCTCT 582
Db 9607 TCTATCGATGTTCAAGATAACCGCAATGTTAGTTTCATGGCTCCTCCCTCATCTCCTCT 9666
QY 583 CCACCTTCACTCTTTGGATCAATTAAGACACTATGATGATTCCTCAAAACAATGTTGGGT 642
Db 9667 CCACCTTCACTCTTTGGATCAATTAAGACACTATGATGATTCCTCAAAACAATGTTGGGT 9726
QY 643 TTTGAAGCAAAATAGTAGTTTCAGGCATTTTCAGGTGATGTTGGTCCAAAGTGAACCAATG 702
Db 9727 TTTGAAGCAAAATAGTAGTTTCAGGCATTTTCAGGTGATGTTGGTCCAAAGTGAACCAATG 9786
QY 703 ATGCTTACATTCGGTGAAGAAGATTTCCTCGTTTCTAAATTCGAATTAAGAAACCAATGAG 762
Db 9787 ATGCTTACATTCGGTGAAGAAGATTTCCTCGTTTCTAAATTCGAATTAAGAAACCAATGAG 9846
QY 763 CTTTCAATGAGCTTCGATCAGATGTTTCTGATGATGCTCGGAGTAAGTCTTTGTGCA 822
Db 9847 CTTTCAATGAGCTTCGATCAGATGTTTCTGATGATGCTCGGAGTAAGTCTTTGTGCA 9906
QY 823 GCTACAAGATTAGCTTCAGAGCAAGCTTCTTCAGCAGCAAAAGACATTTCTAATAACGTT 882
Db 9907 GCTACAAGATTAGCTTCAGAGCAAGCTTCTTCAGCAGCAAAAGACATTTCTAATAACGTT 9966
QY 883 GTTACTCAAGGTTTCTCTCAACTTATTTGGCTCAAAATACCTTCACCTCTGTTCAAGAA 942
Db 9967 GTTACTCAAGGTTTCTCTCAACTTATTTGGCTCAAAATACCTTCACCTCTGTTCAAGAA 10026
QY 943 ATACTATCTCATTTGCCGCATACCTCGCTCGATTATTCATCTCGAG 988
Db 10027 ATACTATCTCATTTGCCGCATACCTCGCTCGATTATTCATCTCGAG 10072

RESULT 11
AF375967
LOCUS AF375967 1182 bp mRNA linear PLN 02-JUN-2003
DEFINITION Lycopersicon esculentum bell-like homeodomain protein 4 (BL4) mRNA, partial cds.
ACCESSION AF375967
VERSION AF375967.1 GI:31323448
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon. 1 (bases 1 to 1182)
REFERENCE Ron N., Parnis A. and Lifschitz, E.
AUTHORS Direct Submission
JOURNAL Submitted (03-MAY-2001) Biology, Technion, Haifa, Israel
FEATURES
source
1..1182
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/db_xref="taxon:4081"
<1..1182
/genes="BL4"
<1..973
/genes="BL4"
/note="TBL4"
/codon_start=2

/product="bell-like homeodomain protein 4"
/protein_id="AAP47026.1"
/db_xref="GI:31323448"
translation="RDQCSIDISGVNTHAFQRRFDSELTSCNSRNLSLSPGSKPV
YLSOFITGSLRYLWMOEILSEIAQLSLQHNLVGYRGNGTENGANTSFALNSDAGRY
AAMSDDSPDLMLGCEAKKNLVALLQVDDQVNOCLDETHWISAFHATYELDPSIH
ARALQITISLYKNLGENKNSHRSNRGTFQORMRKSGREVLKHSFKSSGHFOOL
KRDHQLRFPQGLPERSVSLRAWNFQFLHYFKDAEKQLLAVKSLGSLTRQSUNHF
INARVLWKPMI BEMYAEMNRRKIRAGNEEDHRRNHKILIESHLFTMK"
Query Match 13.5%; Score 255; DB 8; Length 1182;
Best Local Similarity 61.8%; Pred. No. 2.8e-47;
Matches 521; Conservative 0; Mismatches 290; Indels 32; Gaps 5;
QY 826 ACAAGATTAGCTTCAGAGCAAGCTTTTGCAGCAGCAAAACACATTTCTTAATAACGTTGTT 885
Db 59 AGACGCTTTGATTTCAGAGTTAACTTCTTGCACAGCAGAAATCTGCTTTTAAGCTTTGGT 118
QY 886 -----ACTCAAGGTTTCTCTCAACTTATATTGGCTCAAAATACCTTCATCTCTGTT 936
Db 119 TCATACAAACCTGTCTACCTTTTCAAAATTTTAACTGGATCAAGATATCTCCGTGTAATG 178
QY 937 CAAGAAATATCTATCTCAATTTTGGCGCATATCGCTCGATTTATTCATCTCAGAGCAACGAG 996
Db 179 CAAGAAATATCTCTCTGAAATGCTCAGTTATCACTTCAAAATCATAACTTGTGAGGCTAT 238
QY 997 TCAGAGCTGCTA-GTTCAAGCTTTACTTCAAGCTTTTGAGATATTAATCACTGATTTCTGA 1055
Db 239 CGAGAAATGGAGCGAGAAATGGTCAACACCTTCATTTGCTTTGAAATTCAGATGCTGGA 298
QY 1056 TGGTGATTCTTAATAACTCGGAGCGGGTTTCGGATCTACATTTCAAAAGGAGCAATTAGA 1115
Db 299 AGGGCT-----ATGCACCATGAGTTCTGATGATTCACCTGATGGTCTGATGGTTGTGA 354
QY 1116 AGCAAGAAACCATCTCTTGGATCTTCTTCAAAATGGTGGATGATGATGATGATGATGATG 1175
Db 355 AGCTTAAGAAAGAAATCTGTTAGCTTTGTTGCAAGTGGTTGAGCATCAATCAACCAATG 414
QY 1176 CGTAGATGAGATTCATACGCTTATATCAGCGTTTATCAGCTGCTCAACCGAGTTAGATCCACA 1235
Db 415 CTTAGATGAGATTCATATGATGATCTCTGATTTTCATCTGCTGACCGAGTTGATCTAG 474
QY 1236 GTTACACACCGGTTTGCCTTCCAAACCGTTTCTTATATACAAAGAACTGTA-GAGAGA 1294
Db 475 TATACATGCTCGTTTTCGCTTCAACAAATCTCTCTCTGTTACAAAAACCTGAGGGGAGA 534
QY 1295 GAATCTCAAGAGATTAATCTCTATGGATCTGTATTGGAGAGAGCAAGACAAAGACTC 1354
Db 535 GAATAAGCAATCATCTCTCGAATGGAGAAACATTTCAACAAAGATGAGAGAAAGGG 594
QY 1355 AA-----GAAACCTCTATGTTCCACCAGCA-TTGCCCTTCTTCACGAGCT 1397
Db 595 GAGTAGAGAGTCAATTTTGAACATCATTCATTCAAAAGCAGTGGGCACCTTTCAGCAGCT 654
QY 1398 GAAACGAAAGAACCATCAGATTTGGAGACCTCAACGAGGTTTGCTCAGAAATCTGTTTC 1457
Db 655 AAAACGAAAGATCATCAGTTATGGAGGCCACAAAGAGGTTTGCAGAAAGATCTGATC 714
QY 1458 GGTCTTACGGAATTTGGATGTTCCAAAACTTCTTCACTTCACTTCACTTCACTTCACTTCACT 1517
Db 715 GGTCTTACGAGATGATGATTTTCAAACTTCTTCACTTCACTTCACTTCACTTCACTTCACT 774
QY 1518 ACATCTCTAGCTATACGAGTGGCTTGAAGAGTCAAGAGTCAAGTATCAACTGTTTATATA 1577
Db 775 GCAATTTGCTAGCAGTAAAAAGTGGATTGACAAGAGTCAAGTATCGAATTTGTTTCAATA 834
QY 1578 TGCAGGTTAGGCTATGGAGCGCATGATAGAGAGATGTATCGGAAATGAACAAGAG 1637
Db 835 TGCTCGATTCCTCTATGGAACCAATGATAGAGAAATGTATGCTGAGATGAACAGAG 894
QY 1638 GAA 1640
|||

```

Db      895 AAA 897

RESULT 12
AF406703      2065 bp mRNA linear PLN 14-JUL-2003
LOCUS      Solanum tuberosum BEL1-related homeotic protein 30 (Bel30) mRNA,
DEFINITION complete cds.
ACCESSION AF406703
VERSION AF406703.1 GI:22652126
KEYWORDS
SOURCE
ORGANISM
Solanum tuberosum (potato)
Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 2065)
AUTHORS Chen,H., Rosin,F.M., Prat,S. and Hannapel,D.J.
TITLE Interacting transcription factors from the three-amino acid loop
extension superclass regulate tuber formation
JOURNAL Plant Physiol. 132 (3), 1391-1404 (2003)
PUBMED 12857821
REFERENCE 2 (bases 1 to 2065)
AUTHORS Chen,H., Rosin,F.M. and Hannapel,D.J.
TITLE A KNOX protein of potato interacts with several members of the TALE
family of transcription factors
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 2065)
AUTHORS Chen,H., Rosin,F.M. and Hannapel,D.J.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-2001) Department of Horticulture, Iowa State
University, 257 Horticulture Hall, Ames, IA 50011-1100, USA
FEATURES
Source
1. .2065
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Desiree"
/db_xref="taxon:4113"
1. .2065
/gene="Bel30"
39. .1976
/gene="Bel30"
/Note="StBel30"
/codon_start=1
/product="BEL1-related homeotic protein 30"
/protein_id="AA03627.1"
/db_xref="GI:22652127"
/translation="MATYFPSPNNORDADQTFQYFROSLPSYSEASNAPENMMVFN
YSSGAYSDMLTGTSQQOHCIDIPSIGATPNTSQEILSLGSGOMGICDFSWED
SENEMLADNVAQNVQGLSLGSLNSIPSGTGISHVQSNPQGGKFMFGDGV
NSQPEQNAQYFPDPNFRGLDAMKGNYSPTGSSARTIPSSKYLKALQYLDEYV
VONKLAUKSLMLDEIDRYQYHOMQIVVSSFDVAVAGAAKPYTALQTSRHPR
CLRDAICQIPASRLSGEODASENSKAIGISRLRFVDDHRIQORALQQLGMMQOHA
RPQRLGPSSVSLRAWLFEHLHPYKDSKIMLARQTLGTRQVSNWFINARVLW
KPMVEYKEBAGDAKIDNSSSDVAPRLAKDSKVERGELHONAAEFEOYNSGOI
LESKSHZADVMEGASNAETQOSQVGENOTGEPLPANDNCTLFQDAFVQSDRPFSEF
GSFGSGNVLPNGVSLTGLQOGBSNLPMSETHSVVPLRADDMYSTATTNVPETA
EFNCLDSGNRQOPWLLPSAT"

Query Match 7.0%; Score 132; DB 8; Length 2065;
Best Local Similarity 70.2%; Pred. No. 2.4e-19;
Matches 177; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1375 CAGCATTCCTTTCAGCAGCTGAAACGAAGAACCATCAGATTGGAGACCTCAGGA 1434
DB 1167 CAGCAGAGAGCCCTCAGCAGCTTGGTATGATGCAACAAACATCGCTCGAGCCCTCAGAGG 1226
QY 1435 GGTTCGCTGAGAAATCTGTTTCGGTTCTACGGAATGGATGTTCCAAAACCTCTCTTCA 1494
DB 1227 GGATTGCTGGAAGCTCTGTTTCAGTTTGGTGTCTTGGCTCTTGGACACTTCTTCAT 1286

```

```

QY 1495 CCTTACCCGAAAGATTGGGAGAACATCTTCTAGCTATACGAGTGGCTTGACNAGACT 1554
DB 1287 CCTTACCCGAAAGATTCTGACAAAATATTGCTAGCAAGGCAAACTGGCTTAACGAGACT 1346
QY 1555 CAGGTATCAAACTGGTTTATAAATGCGCGGTGTAGGCTATGGAAGCCGATGATAGAAGAG 1614
DB 1347 CAGGTATCAAACTGGTTTATAAATGCAACCGGTGCGTCTTTGGAAACCCATGTTGAGGAA 1406
QY 1615 ATGTATGCGGAA 1626
DB 1407 ATGTACAAAGAA 1418

RESULT 13
AF406697      2735 bp mRNA linear PLN 14-JUL-2003
LOCUS      Solanum tuberosum BEL1-related homeotic protein 5 (Bel5) mRNA,
DEFINITION complete cds.
ACCESSION AF406697
VERSION AF406697.1 GI:22652114
KEYWORDS
SOURCE
ORGANISM
Solanum tuberosum (potato)
Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 2735)
AUTHORS Chen,H., Rosin,F.M., Prat,S. and Hannapel,D.J.
TITLE Interacting transcription factors from the three-amino acid loop
extension superclass regulate tuber formation
JOURNAL Plant Physiol. 132 (3), 1391-1404 (2003)
PUBMED 12857821
REFERENCE 2 (bases 1 to 2735)
AUTHORS Chen,H., Rosin,F.M. and Hannapel,D.J.
TITLE A KNOX protein of potato interacts with several members of the TALE
family of transcription factors
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 2735)
AUTHORS Chen,H., Rosin,F.M. and Hannapel,D.J.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-2001) Department of Horticulture, Iowa State
University, 257 Horticulture Hall, Ames, IA 50011-1100, USA
FEATURES
Source
1. .2735
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Desiree"
/db_xref="taxon:4113"
1. .2735
/gene="Bel5"
148. .2214
/gene="Bel5"
/Note="StBel5"
/codon_start=1
/product="BEL1-related homeotic protein 5"
/protein_id="AA03621.1"
/db_xref="GI:22652115"
/translation="WYQGTSDNTNIQADHQORHNHNSNNNTIOTLYLMPNNYMQG
YTTSDTQQOQLFLNSSFPAASNALCHANIQAPLQOQHFGVGPULPAVSLHDDQINHHG
LIQRMNNQDQQQVIVPSFSGVTSATCGGTTDLASQAFORPITPCHLQOQQQQG
GLSLSPQLQOQISFNINISSSPRTNVTIRGTLGSSNMVLGSKYLKAAQLLD
EVNIVGKSIKGDQKDNMKNMSPPLASDVNTNNSGGSSSRKQNEVAELTTAQ
ROELMKKALMLAEVEQRYVHOMQIIVLSEFQVAGVAGSAKSYQALHAIK
QFRCFKDAIEQVATSKLGEELGGKIEGSLKRVDFVHLRQORALQQLGMMQPN
MRPQGLPERAVSLRAWLFEHLHPYKDSKIMLARQTLGTRQVSNWFINARVL
WKPWVEYKEBAGDAKIDNSSSDVAPRLAKDSKVERGELHONAAEFEOYNSGOI
ABISTSTISPTGASLHAHNFSLGFSFNNMDNTTTVDHIENNAKKORNMKFSF
SSILSSVMEAKARESSNKGFTNPLMAAYAMGDFQFDPHQQMTANFHNGVSLT
GLPFSNENLAMPVSQQNYLNDLSRSEMGSHYMRMGYENIDFQSGNKEFTPTQLLPDFV
TGNLGT"

Query Match 6.8%; Score 128.6; DB 8; Length 2735;

```

ORIGIN

Best Local Similarity 66.3%; Pred. No. 1.4e-18;		Matches 185; Conservative 0; Mismatches 94; Indels 0; Gaps 0;	
Qy	1375	CAGCATTGCTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACCTCAACGA	1434
Db	1279	CAACAAACGCGCTGCAACAGCTAGGAATGATCAACCAATGCTTTGGAGACCCCAAGA	1326
Qy	1435	GGTTTGCTGAGAAATCTGTTTGGTTCTACGGAATGGATGTTCCAAAATCTTCCTCAC	1494
Db	1327	GGTTTACCTGAAAGAGCTGCTCTGCTTCTGCTTGGCTTTTCGAGCAATTTCTTCAT	1386
Qy	1495	CTTTACCCGAAAGATTGCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAGAAGT	1554
Db	1387	CTTTACCCGAAAGACTCAGACAAAATCATCTTGTCTAAGCAACCGGGCTTAACAAGGAGC	1446
Qy	1555	CAGGTATCAAACTGGTTTATAATGCGCGGTTAGCTTGAAGCGCGATAGAGAG	1614
Db	1447	CAGCTCTCACTAAGGTTTATTAATGCTGAGTTCGATTATGGAAGCCCAATGGTTGAAGAA	1506
Qy	1615	ATGTATCGGAAATGAAACAAAGAGGAGCTCAATAACAGT	1653
Db	1507	ATGTACTTGAAGAGTGAAGATCAAGAACAAACAGT	1545
RESULT 14			
AF375966			
Lycopersicon esculentum bell-like homeodomain protein 2 (BL2) mRNA, linear PLN 02-JUN-2003			
complete cds.			
ACCESSION AF375966			
VERSION AF375966.1 GI:31323446			
KEYWORDS Lycopersicon esculentum (tomato)			
SOURCE Lycopersicon esculentum			
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.			
1 (bases 1 to 2636)			
REFERENCE Ron.N. Parnis.A. and Lifschitz,E.			
AUTHORS Direct Submission			
TITLE Submitted (03-MAY-2001) Biology, Technion, Haifa, Israel			
JOURNAL Location/Qualifiers			
FEATURES			
1..2636			
/organism="Lycopersicon esculentum"			
/mol_type="mRNA"			
/db_xref="taxon:4081"			
1..2636			
/gene="BL2"			
100..2199			
/genes="BL2"			
/note="TBL2"			
/codon_start=1			
/product="bell-like homeodomain protein 2"			
1			
/protein_id="AAP47025.1"			
/db_xref="GI:31323447"			
/translation="MYQGTSDNNIQAHHQOQHNNGNNSNNIQTLYLMPNSYMQG			
YTTDTQHQQQQHQHLLFNSAPAGNALSHANIQHAPLQQHFGVPLPVSLLH			
DOINHHLLQRMWNNQSQQVIVPSSTVVSATGCGTTDLASQAFQPIVSPPT			
YHQCOQOQGLSLSPQOQOQISFNNNISSSPRTNNVTIRGTWDCSSNMILGSK			
YKQAQELLDEVNIVGSKNGDKDKNSMKEILPLVSDVNTNSGGGGESSRSO			
KNEVALTTAQOELQAKKALLMLEVEQRYQYHHOMLIIVSSPEQVAGVSAX			
SYQLALHAISQKFCRLKDAISEQVKATSKLGEDEGLGKTEGSLKFLFVHHLRQR			
ALQQLGMQPNAPRQGLPERAVSLRAWLFHLHPYKPDSDKIMLAKQGLTRSQ			
VSNFNFARVRLKPMVEEMYLEEVKNQNSNTSGDNKNTINISAPNEKQPIIT			
SLSLQDGTQALISTYSTISPTAGASHAHNFPLGSGFNNTTVDHLENNAKK			
PRNDHMFSPSSILSSVMEAKARESTNKGFTNPLMAAYAGDFRFDPHDQMTAN			
FHNNGVSLTGLPPEENLAMPVSQNYLSNLSGSRPEIGSHYNGVENDIDFQSGNK			
RPTQLLPDFVTNGUT"			
ORIGIN			
Query Match 6.7%; Score 127; DB 8; Length 2636;			
Best Local Similarity 65.9%; Pred. No. 3.3e-18;			
Matches 184; Conservative 0; Mismatches 95; Indels 0; Gaps 0;			
Qy	1375	CAGCATTGCTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACCTCAACGA	1434
Db	1279	CAACAAACGCGCTGCAACAGCTAGGAATGATCAACCAATGCTTTGGAGACCCCAAGA	1326
Qy	1435	GGTTTGCTGAGAAATCTGTTTGGTTCTACGGAATGGATGTTCCAAAATCTTCCTCAC	1494
Db	1327	GGTTTACCTGAAAGAGCTGCTCTGCTTCTGCTTGGCTTTTCGAGCAATTTCTTCAT	1386
Qy	1495	CTTTACCCGAAAGATTGCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAGAAGT	1554
Db	1387	CTTTACCCGAAAGACTCAGACAAAATCATCTTGTCTAAGCAACCGGGCTTAACAAGGAGC	1446
Qy	1555	CAGGTATCAAACTGGTTTATAATGCGCGGTTAGCTTGAAGCGCGATAGAGAG	1614
Db	1447	CAGCTCTCACTAAGGTTTATTAATGCTGAGTTCGATTATGGAAGCCCAATGGTTGAAGAA	1506
Qy	1615	ATGTATCGGAAATGAAACAAAGAGGAGCTCAATAACAGT	1653
Db	1507	ATGTACTTGAAGAGTGAAGATCAAGAACAAACAGT	1545
RESULT 15			
AK067939			
Oryza sativa (japonica cultivar-group) cDNA clone:J013124I05, full			
insert sequence.			
ACCESSION AK067939			
VERSION AK067939.1 GI:32977957			
KEYWORDS FLJ CDNA; CAP trapper.			
SOURCE Oryza sativa (japonica cultivar-group)			
ORGANISM Oryza sativa (japonica cultivar-group)			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
Erihartoideae; Oryzeae; Oryza.			
1			
REFERENCE The Rice Full-length cDNA Consortium, National Institute of			
AUTHORS Agrobiological Sciences Rice Full-length cDNA Project Team,			
Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,			
Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,			
Ojima,K., Namiki,T., Ohmeda,E., Yahagi,W., Suzuki,K., Li,C.,			
Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN,			
Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,			
Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,			
Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Oota,N., Ota,Y.,			
Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,			
Yoshino,M. and Hayashizaki,Y.			
Collection, mapping, and annotation of over 28,000 cDNA clones from			
japonica rice			
Science 301 (5631), 376-379 (2003)			
MEDLINE 22752273			
PUBMED 12869764			
REFERENCE 2 (bases 1 to 2570)			
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,			
Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,			
Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hirooka,T.,			
Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,K., Imamura,K.,			
Imotani,K., Ishibiki,S., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,			
Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,			
Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,			
Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,			
Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,			
Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,			
Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,			
Namiki,T., Nariikawa,R., Niikura,J., Nishi,K., Nomura,K.,			
Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,			
Osato,N., Ota,Y., Ootomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,			

Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
Sugiyama, A., Suzuki, K., Tagami, M., Tanaka, T., Tanaka, T., Tomaru, A.,
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and
Yoshimura, A.
Direct Submission
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression; 2-1-2 Kanondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp,
Tel: 81-29-838-7007, Fax: 81-29-838-7007)
This clone is one of the 28K full-length cDNA clones from japonica
rice.

URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,
Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and
Yamamoto, M.

PAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y.,
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,
Mizuno, K., Nariawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,
Sugiyama, A., Suzuki, F., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
Yoshimura, A., Matsubara, K. and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanacaki, T.,
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,
Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A. and Hayashizaki, Y.

FEATURES
source
1. 2570
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J013124105"

ORIGIN

Query Match 6.7%; Score 126.4; DB 8; Length 2570;
Best Local Similarity 68.4%; Pred. No. 4.5e-18;
Matches 175; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 1375 CAGCATTCCTTCTTCACAGCTGAACGAAAGAACCATCAGATTGGAGACCTCAACGA 1434
DB 1482 CAACAGCGCGTTTCCAGCAGTATGGTTTGTACAGCAAAATGCTTGGAGGCCACAGAGG 1541
QY 1435 GGTTCCTGAGAAATCTTCGTTCTTACGGAATGGATGTTCCAAACTTCCTTCAC 1494
DB 1542 GGACTGCCGGAATCTAGTTTCATCTTCGTTGGCTTTTGAACACTTCCTTCAC 1601
QY 1495 CTTACCCGAAGATTCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAGAAGT 1554
DB 1602 CCGTATCCAAAGATTCAGAAAGCTGATGCTAGCGACAACTGGCTTAACAAGAAGT 1661
QY 1555 CAGGTATCAACTGTTTATTAATCCGCGGTTAGGCTATCGAAGCCGATGATAGAGAG 1614
DB 1662 CAGATTTCAATTTGTTTCATTAATGCCCGTGTCCCGCTGTGGAAACCGATGATCGAAGAC 1721
QY 1615 ATGTATCGGAATGA 1630
DB 1722 ATGTATAAGAAGAGA 1737

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 2, 2004, 08:25:17 ; Search time 113.5 Seconds
(without alignments)
9390.037 Million cell updates/sec

Title: US-09-423-575-1
Perfect score: 3284
Sequence: 1 attaggtataaaatgttc.....ttaaaaaaaaaaaaaaaaaa 1886

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09423575/runat.01092004.161105.18548/app.query.fasta_1.2055
-DB=A_Geneseq.25Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOPCL=0
-LCOEPT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09423575@CGN.1.145@runat.01092004.161105.18548 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq.29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2484	75.6	473	5 AAU76514	Arabidops
2	2484	75.6	473	5 AAU76885	Arabidops
3	472	14.4	680	6 ADB23150	Enviro
4	472	14.4	680	7 ADD30304	Plant yle
5	444	13.5	514	3 AAG14898	Arabidops
6	444	13.5	524	3 AAG14897	Arabidops
7	444	13.5	524	4 AA502524	Arabidops
8	444	13.5	533	3 AAG14896	Arabidops
9	424	12.9	739	3 AA502510	Arabidops
10	424	12.9	739	7 ADE37297	Plant yle

11	419	12.8	459	3 AAG14140	Arabidops
12	419	12.8	459	3 AAG32446	Arabidops
13	419	12.8	482	3 AAG32445	Arabidops
14	419	12.8	482	3 AAG32445	Arabidops
15	419	12.8	491	3 AAG21086	Arabidops
16	419	12.8	539	3 AAG21085	Arabidops
17	415	12.6	418	3 AAG21087	Arabidops
18	412.5	12.6	610	4 AAE02469	Arabidops
19	412.5	12.6	610	7 ADE37117	Plant yle
20	411.5	12.5	423	3 AAG32447	Arabidops
21	411.5	12.5	423	3 AAG32447	Arabidops
22	408	12.4	627	4 AAE02525	Arabidops
23	408	12.4	627	7 ADC46619	Thalecres
24	408	12.4	627	7 ADD30787	Plant yle
25	403	12.3	471	3 AAG50499	Arabidops
26	403	12.3	494	3 AAG50498	Arabidops
27	395.5	12.0	435	3 AAG50500	Arabidops
28	379	11.5	575	3 AAG21938	Arabidops
29	379	11.5	598	3 AAG21937	Arabidops
30	377	11.5	608	5 ABB79411	Hordeum v
31	376.5	11.5	611	5 ABB79412	Hordeum v
32	348.5	10.6	500	3 AAG21939	Arabidops
33	348.5	10.6	612	6 AAO27114	QSH-1 gen
34	348.5	10.6	612	6 AAO27115	QSH-1 gen
35	338	10.3	153	3 AAB32901	Pinus rad
36	332	10.1	176	3 AAB32619	Eucalyptu
37	327.5	10.0	107	3 AAB32960	Pinus rad
38	326	9.9	126	3 AAB32644	Eucalyptu
39	301	9.2	233	5 ABB60937	Novel flo
40	299	9.1	202	3 AAG08089	Arabidops
41	243	7.4	293	3 AAG50529	Arabidops
42	243	7.4	316	3 AAG50528	Arabidops
43	235.5	7.2	257	3 AAG50530	Arabidops
44	232	7.1	126	3 AAB32653	Eucalyptu
45	222	6.8	86	3 AAB32893	Pinus rad

ALIGNMENTS

RESULT 1
AAU76514
ID AAU76514 standard; protein; 473 AA.
XX AAU76514;
AC AAU76514;

DT 05-JUN-2002 (first entry)

DE Arabidopsis Athl protein.

KW Grass; plant; herbicide resistance; grass flowering inhibitor; baseball;
KW cricket; football; golf; rugby; soccer; tennis; lawn; park;
KW athletic field; animal feedstuff; Athl protein.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN WO200214524-A2.

XX 21-FEB-2002.

XX 16-AUG-2001; 2001WO-EF009572.

XX 18-AUG-2000; 2000US-0226422P.

XX 27-NOV-2000; 2000US-0253274P.

XX 27-NOV-2000; 2000US-0253327P.

XX 22-JUN-2001; 2001US-0300220P.

(ADVA-) ADVANTA SEEDS BV.

XX Van Der Valk P, Van Dun CMP, Smeekens SCM, Proveniers MCG;

XX WPI; 2002-257611/30.

XX N-FSDS; ABK11039.

PT New genetically modified grasses that express inhibited generative
PT propagation, or herbicide resistance, useful for forage (e.g. cattle
PT feedstuff) or amenity purposes (e.g. for use in an athletic field, lawn
or park).

XX Example 1; Fig 6; 56pp; English.

XX The invention relates to a grass plant, which has been genetically
CC modified to substantially inhibit generative propagation and carry
CC herbicide resistance. The grass is useful for growth and/or propagation
CC of grasses. The grass is used to play at least one sport (e.g. baseball,
CC cricket, football, golf, rugby, soccer or tennis), or used at least in a
CC portion of an athletic field, lawn or park. The grass is fed to animal
CC (e.g. cattle, goat, horse or sheep) or used as an animal feedstuff. The
CC present sequence represents the amino acid sequence of Arabidopsis Ath1
CC protein, which was used to inhibit flowering in grasses

XX Sequence 473 AA;

Alignment Scores:
Pred. No.: 5,45e-250 Length: 473
Score: 2484.00 Matches: 473
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 75.64% Indels: 0
DB: 5 Gaps: 0

US-09-423-575-1 (1-1886) x AAU76514 (1-473)

QY 304 ATGGCAACAAACAAACAAACAACTTTTCTCGATAATGTCAGCTAACCAA 363
DB 1 MetAspAsnAsnAsnAsnAsnAsnThrPheSerSerLeuAspAsnValMetThrAsnGln 20
QY 364 AATCTCTCTCATGGATTTATACCTTCAAGAGAGATTCACTTCATTTCTCAACATG 423
DB 21 AsnProLeuLeuMetAspPheLeuProSerArgGluAspSerThrSerPheSerThrMet 40
QY 424 CTTCATGGAATACCATCAGATCCTCTCAAAATGGGTGGCTTTGATATTTCAAT 483
DB 41 LeuProTrpAsnThrIleArgSerAspProLeuGlnMetGlyGlyPheAspIlePheAsn 60
QY 484 TCTATGCTGACTAAACAATACTTATCATCTCTCCACGGTCTATCGATGTCAGATAAC 543
DB 61 SerMetLeuThrAsnLysTyLeuSerSerSerProArgSerIleAspValGlnAspAsn 80
QY 544 CCAATGTTGAGTTTCATGGCTCCTCTCTCATCTCTCCACATTCATCTTTCGATCAT 603
DB 81 ArgAsnValGluPheMetAlaProProProHisProProLeuHisProLeuAspHis 100
QY 604 TTAAGACACTATGATGATTCCTCAACCAACATGCGGGTTTGAAGCAATAGTGAATTT 663
DB 101 LeuArgHisTyAspAspSerSerAsnAsnMetTrpGlyPheGluAlaAsnSerGluPhe 120
QY 664 CAGGCATTTTCAGGTGATGTTGGTCCCAAGTGAACCAATGATGTTTACATTCGGTGAAGAA 723
DB 121 GlnAlaPheSerGlyValValGlyProSerGluProMetMetSerThrPheGlyGluGlu 140
QY 724 GATTTCCGGTTTCTTAATTCGAATAAAGAAACAATAGCTTTTCATGAGCTTTGCAATCA 783
DB 141 AspPheProPheLeuLeuSerAsnLysArgAsnAsnGluLeuSerLeuSerLeuAlaSer 160
QY 784 GATGTTCTTGATGATGTCGGAGATAGTCTTTGTCAGCTACAAGATAGCTTCAGAG 843
DB 161 AspValSerAspGlyCysSerGluLeuSerLeuCysAlaAlaThrArgLeuAlaSerGlu 180
QY 844 CAAGTCTTTCAGCAGCAAGACATTTCTTAATACGTTGTACTCAAGGTTTCTCTCAA 903
DB 181 GlnAlaSerCysSerSerLysAspIleSerAsnAsnValValThrGlnGlyPheSerGln 200
QY 904 CTTATATTGGCTCAAAATACCTTCACCTCTTCAAGAAATACTATCTCATTTCCGCGCA 963
DB 201 LeuIlePheGlySerLysTyLeuHisSerValGlnGluIleLeuSerHisPheAlaAla 220

QY 964 TACTCCTCGATTATTCATCTCGAGAACCGAGTCAGGAGCTGCTAGTTCAGCCTTTACT 1023
DB 221 TyrSerLeuAspTrpSerSerArgGlyThrGluSerGlyAlaAlaSerSerAlaPheThr 240
QY 1024 TCACGTTTTGAGAAATAAATAGTTTCTTGATGGTGATTCTTAATAACTCGAGGCGGT 1083
DB 241 SerArgPheGluAsnIleThrGluPheLeuAspGlyAspSerAsnAsnSerGluAlaGly 260
QY 1084 TTCGGATCTACATTTCAAAGGAGAGCATTAGAGCAAGAAACCAACCATCTTGGATCTT 1143
DB 261 PheGlySerThrPheGlnArgAlaLeuGluAlaLysLysThrHisLeuLeuAspLeu 280
QY 1144 CTTCAATGGTGGATGATCGATATAGTCTATCGGTAGATGAGATTCATACGGTTTATCA 1203
DB 281 LeuGlnMetValAspAspArgTrpSerHisCysValAspGluIleHisThrValIleSer 300
QY 1204 GCCTTCCATGCTGCAACCGAGTTAGATCCACAGTTTACACACCCGGTTTGCCTCCAAACC 1263
DB 301 AlaPheHisAlaAlaThrGluLeuAspProGlnLeuHisThrArgPheAlaLeuGlnThr 320
QY 1264 GTTTCCTTCTTATACAGAACCTGAGAGAGAGAAATCTCCAGAGATTAATCTCTATGGA 1323
DB 321 ValSerPheLeuTyTrpLysAsnLeuArgGluArgIleCysLysLysIleIleSerMetGly 340
QY 1324 TCTGTATTGAGAGAGGCAAGCAAGACTCAAGAAACCTCTATCTTCCACCGCATTCG 1383
DB 341 SerValLeuGluArgGlyLysAspLysThrGlnLeuThrSerMetPheHisGlnHisCys 360
QY 1384 CTTCTTCAGAGCTGAACGAAGAACCATCAGATTGGAGACCTCAACGAGTTTGGCT 1443
DB 361 LeuLeuGlnGlnLeuLysArgLysAsnHisGlnIleTrpArgProGlnArgGlyLeuPro 380
QY 1444 GAGAAATCTGTTTCGGTCTCTCGGAATGGATGTTCCAAACCTCTTCCACCTTACCAG 1503
DB 381 GluLysSerValSerValLeuArgAsnTrpMetPheGlnAsnPheLeuHisProTyTrp 400
QY 1504 AAAGATTTCGAGAAACATCTTCTAGCTATACGAATGGCTTGACAAAGTCAGATATCA 1563
DB 401 LysAspSerGluLysHisLeuAlaIleArgSerGlyLeuThrArgSerGlnValSer 420
QY 1564 ACTGCTTTTATAAATCCGCGGTAGGCTATCGAGCCGATGATAGAGAGATGATGCG 1623
DB 421 AsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluGluMetTyAla 440
QY 1624 GAAATGAACAGAGGAAGCTCAATAACAGTCACTTCAACCCCAACGAGCAACTCTTCA 1683
DB 441 GluMetAsnLysArgLysLeuAsnAsnSerHisIleGlnProAsnGlyProThrLeuArg 460
QY 1684 ATGCCAAATCTGTTATGATGAGCCCAAGCAATGCATAA 1722
DB 461 MetProLysSerValMetMetSerGlnAlaMetHisLys 473

RESULT 2

AAU76885
ID AAU76885 standard; protein; 473 AA.

AC AAU76885;

DT 05-JUN-2002 (first entry)

DE Arabidopsis thaliana homeobox gene Ath1 encoded protein.

XX Grass; plant; transgenic; flowering inhibition; inflorescence;
KW tiller production; delayed heading; gibberellic acid; phytohormone; golf;
KW genetically modified grass; athletic field; sport; baseball; cricket;
KW football; rugby; soccer; tennis; lawn; landscaping; cattle; horse; sheep;
KW goat; animal feedstuff; Ath1; homeobox gene.

OS Arabidopsis thaliana.

XX WO200214486-A2.

XX 21-FEB-2002.

PD

QY 1684 ATGCCAAAATCTGTTATGATGAGCCAAGCAATGCATAAA 1722
|||
Db 461 MetProLysSerValMetMetSerGlnAlaMetHisLys 473
|||

RESULT 3

ADB23150
ID ADB23150 standard; protein: 680 AA.

Qy	514	TC	CC	ACGGTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTCATGGCTCTCTCTCTCTCT	573
Db	91	Gl	uA	aAaSerilethrAlaAaAspAsnIleSerValLeuHisGlyIyrProProArg	110
Qy	574	CAT	CC	CTCCTCCACTTCATCCCTTTGGATCATTTAAAGACACTATGATTCCTCAAAACAAC	633
Db	111	Val	Gln	TyrSerLeuIyrGly-----SerHisGlnValAspProthrHisGln	126
Qy	634	AT	GT	GGGGTTTTTGAAGCAATAGTAGTTTCAGGCAATTTTCAGGTGTAGTGTGCTCAAGT	693
Db	127	Gln	Ala	aAaCysGlnThrProArgAlaGlnGlnGlyLeuSerLeuThrLeuSerSerGln	146
Qy	694	GA	CC	CAATGATGCTACATTCGCTGAAGAAGATTTCCCGTTCTTAATTCGATATAAAGA	753
Db	147	Gln	Gln	GlnGlnHisHisGlnGlnHisGlnProIleHisValGlyPheGlySer	166
Qy	754	AA	CA	ATGAGCTTTTCATTCAGTCTGTGATCAGATGTTTCTGATCAATGCTCGGAGATAAGT	813
Db	167	Gly	His	GlyGluAspIleArgValGlySer-----	176
Qy	814	CT	TT	GTGCAGCTACAAGATTAGCCTCAGACAAGCTTCTTGACAGCAAGAACATTTCT	873
Db	177	-----	-----	-----GlySerThr	179
Qy	874	AA	TA	ACGTTTACTCAAGGTTTCTCTCAACTTATATTTGGCTCAAAATACCTTTCACTCT	933
Db	180	Gly	Ser	GlyValThrSrnglyIleAlaAsnLeuVal--SerSerIysIyrLeuIysAla	198
Qy	934	GT	TC	CAGAATACTACTCATTTCCGCGCATACTCGCTCGATTATTCATCTCGAGGAACC	993
Db	199	Ala	Gln	LeuLeuAspGluValValAsnAlaAspSerAsp-----	212
Qy	994	GAG	TC	AGGAGCTCTAGTTTCAAGCTTTTACTTTCACGTTTTTGAGAATATA--ACTGAGTTT	1050
Db	213	Asp	Met	AsnAlaIysSerGlnLeuPheSerSerIysGlySerCysGlyAsnAspLys	232
Qy	1051	CT	TG	ATGTTGATTAATCAACTCGGAGCGGGTTTCGGATCT-----	1092
Db	233	Pro	Val	GlyGluSerSerAlaGlyAlaGlyGlyGluGlySerGlyGlyAlaGluAla	252
Qy	1093	-----	-----	-----ACATTCAAGAGAGCATTAGAAGCAAGAA	1125
Db	253	Ala	Gly	IyrArgProValGluLeuGlyThrAlaGluArgGlnGlnIleGlnMetIysLys	272
Qy	1126	AC	CA	CTCTCTTGATCTTCTTCAAATGGTGGATGATCGATATGATTCGCTAGATGAG	1185
Db	273	Ala	Iys	LeuSerSerAsnMetLeuHisGluValGluGlnArgIyrArgGlnIyrHisGlnGln	292
Qy	1186	AT	TC	ATACGGTTATATCAGCGTTTCCATGCTGCACCGAGTTA--GATCCACAGTTACAC	1242
Db	293	Met	Gln	MetValIleSerSerPheGluGlnAlaAlaGlyIleGlySerAlaIysSerIyr	312
Qy	1243	AC	CG	GTTTGGCTCCAAACCGTTCTCTTATACAGAACTCGAGAGAGAAATCTGC	1302
Db	313	Thr	Ser	LeuAlaLeuLysThrIleSerArgGlnPheArgCysLeuLysGluAlaIleAla	332
Qy	1303	AA	GA	GATTAATCTCTATGGATCTGTATTGGAGAGAGGCAAAACAGACTCAAGAAACC	1362
Db	333	-----	-----	-----GlyGlnIleIysAlaAlaAsnLysSerLeuGlyGluAsp	346
Qy	1363	TC	TATG	-----TTCCACGAGATTCGCTT	1386
Db	347	Ser	Val	IserGlyValGlyArgPheGluGlySerArgLeuIyrPheValAspHisIleSer	366
Qy	1387	CT	T	CAGCAG-----CTGAAACGAAAGAACCATCAGATTGG	1422
Db	367	Arg	Gln	GlnArgAlaLeuGlnGlnLeuGlyMetIleGlnHisProSerAsnAsnAlaIrp	386
Qy	1423	AG	AC	CTCAGCGTTTGCCTCAGAAATCTGTTTCGGTCTCAGGAAATGGATGTTCCAA	1482
Db	387	Arg	Pro	GlnArgGlyLeuProGluArgAlaValSerValLeuArgAlaIrpLeuPheGlu	406
Qy	1483	AA	CT	CTCTTCCACCTTACCACGAAAGATTCGGAAACATCTTCTAGCTATACGAATGCC	1542

407	HisPheLeuHisProTyrProIysAspSerAspLysHisMetLeuAlaLysGlnThrGly	426
1543	TTGACAAAGAAGTCAGGTATCAAACTGGTTTATAAAATGCGCGGGTTAGGCTATCGAAAGCCG	1602
427	LeuThrArgSerGlnValSerAsnTrpPheIleAsnAlaArgValargLeuTrpLysPro	446
1603	ATGATAGAAGAGATGTATGCGGAATAACACAGAGGAAGCTCAATTAACAGTCACATTCAA	1662
447	MetValGluGluMetTyrMetGluGluMetLysGluGlnAlaLysAsnMetGlySerMet	466
1663	CCCAACGGGACCA-----ACTCTTCGAATG	1686
467	GluLysThrProLeuAspGlnSerAsnGluAspSerAlaSerLysSerThrSerAsnGln	486
1687	CCAAATCTGTTATGATGAGCCAGCAATGCATAAATAAGACAACAAT	1734
487	GluLysSerProMetAlaAspThrAsnTyrHisMetAsnProAsnHis	502
RESULT 4		
ADD30304		
ID	ADD30304 standard; protein; 680 AA.	
XX	ADD30304;	
XX	15-JAN-2004 (first entry)	
DE	Plant yield-related protein from clone G1589.	
XX	transcription factor; transgenic plant; growth rate; senescence;	
KW	seed germination rate; plant vigor; seedling vigor.	
OS	Arabidopsis thaliana.	
XX	WO2003013227-A2.	
PN	20-FEB-2003.	
PD	09-AUG-2002; 2002WO-US025805.	
XX	09-AUG-2001; 2001US-0310847P.	
PR	19-NOV-2001; 2001US-0336849P.	
PR	11-DEC-2001; 2001US-0338692P.	
PR	14-JUN-2002; 2002US-00171468.	
XX	(MEND-) MENDEL BIOTECHNOLOGY INC.	
PA	Patcliffe O, Riechmann JL, Adam LJ, Dubell AT, Heard JE;	
PI	Pilgrim ML, Jiang C, Reuber TL, Creelman RA, Pineda O, Yu G;	
PI	Brown PE;	
XX	WPI; 2003-248221/24.	
DR	N-PSDB; ADD30303.	
XX	New plant transcription factor polynucleotides and polypeptides, useful	
PT	in producing transgenic plants with commercially valuable properties, useful	
PT	such as an alteration in a plant growth characteristic, e.g. growth rate	
PT	or apomixis.	
XX	Disclosure; SEQ ID NO 333; 454pp; English.	
PS	The invention relates to a number of isolated Arabidopsis thaliana cDNA	
XX	sequences and their encoded proteins which are especially transcription	
CC	factor related cDNA's and proteins. The isolated or recombinant plant	
CC	transcription factor polynucleotides and polypeptides are useful in	
CC	producing transgenic plants with commercially valuable properties, i.e.	
CC	modified or altered desirable traits as compared to a reference plant,	
CC	such as an alteration in a plant growth characteristic, e.g. growth rate,	
CC	germination rate of seeds, vigor of plants and seedlings, or leaf and	
CC	flower senescence. Sequence information related to the polynucleotides	
CC	and polypeptides can also be used in bioinformatic search methods. The	
CC	transgenic plant is useful for growing a progeny plant from a parent	
CC	plant. This sequence represents one of the proteins of the invention.	

PR	16-JUL-1999	99US-0144085P.	PR	07-OCT-1999	99US-0158029P.
PR	16-JUL-1999	99US-0144086P.	PR	08-OCT-1999	99US-0158232P.
PR	19-JUL-1999	99US-0144325P.	PR	12-OCT-1999	99US-0158369P.
PR	19-JUL-1999	99US-0144331P.	PR	13-OCT-1999	99US-0159293P.
PR	19-JUL-1999	99US-0144332P.	PR	13-OCT-1999	99US-0159294P.
PR	19-JUL-1999	99US-0144333P.	PR	13-OCT-1999	99US-0159295P.
PR	19-JUL-1999	99US-0144334P.	PR	14-OCT-1999	99US-0159329P.
PR	19-JUL-1999	99US-0144335P.	PR	14-OCT-1999	99US-0159330P.
PR	20-JUL-1999	99US-0144352P.	PR	14-OCT-1999	99US-0159333P.
PR	20-JUL-1999	99US-0144632P.	PR	14-OCT-1999	99US-0159637P.
PR	20-JUL-1999	99US-0144884P.	PR	18-OCT-1999	99US-0159638P.
PR	21-JUL-1999	99US-0144814P.	PR	18-OCT-1999	99US-0159584P.
PR	21-JUL-1999	99US-0145086P.	PR	21-OCT-1999	99US-0160741P.
PR	21-JUL-1999	99US-0145088P.	PR	21-OCT-1999	99US-0160767P.
PR	22-JUL-1999	99US-0145087P.	PR	21-OCT-1999	99US-0160768P.
PR	22-JUL-1999	99US-0145089P.	PR	21-OCT-1999	99US-0160770P.
PR	22-JUL-1999	99US-0145133P.	PR	21-OCT-1999	99US-0160814P.
PR	22-JUL-1999	99US-0145192P.	PR	21-OCT-1999	99US-0160815P.
PR	22-JUL-1999	99US-0145145P.	PR	22-OCT-1999	99US-0160980P.
PR	23-JUL-1999	99US-0145218P.	PR	22-OCT-1999	99US-0160981P.
PR	23-JUL-1999	99US-0145224P.	PR	22-OCT-1999	99US-0160989P.
PR	26-JUL-1999	99US-0145276P.	PR	25-OCT-1999	99US-0161404P.
PR	27-JUL-1999	99US-0145918P.	PR	25-OCT-1999	99US-0161405P.
PR	27-JUL-1999	99US-0145918P.	PR	25-OCT-1999	99US-0161406P.
PR	27-JUL-1999	99US-0145918P.	PR	25-OCT-1999	99US-0161359P.
PR	27-JUL-1999	99US-0145919P.	PR	25-OCT-1999	99US-0161360P.
PR	28-JUL-1999	99US-0145951P.	PR	26-OCT-1999	99US-0161360P.
PR	02-AUG-1999	99US-0146386P.	PR	26-OCT-1999	99US-0161361P.
PR	02-AUG-1999	99US-0146388P.	PR	26-OCT-1999	99US-0161362P.
PR	02-AUG-1999	99US-0146389P.	PR	28-OCT-1999	99US-0161992P.
PR	03-AUG-1999	99US-0147038P.	PR	28-OCT-1999	99US-0161993P.
PR	04-AUG-1999	99US-0147204P.	PR	29-OCT-1999	99US-0162142P.
PR	04-AUG-1999	99US-0147302P.			
PR	05-AUG-1999	99US-0147192P.			
PR	05-AUG-1999	99US-0147260P.			
PR	06-AUG-1999	99US-0147303P.			
PR	06-AUG-1999	99US-0147416P.			
PR	09-AUG-1999	99US-0147493P.			
PR	09-AUG-1999	99US-0147935P.			
PR	10-AUG-1999	99US-0148171P.			
PR	11-AUG-1999	99US-0148319P.			
PR	12-AUG-1999	99US-0148341P.			
PR	13-AUG-1999	99US-0148665P.			
PR	13-AUG-1999	99US-0148684P.			
PR	16-AUG-1999	99US-0149368P.			
PR	17-AUG-1999	99US-0149175P.			
PR	18-AUG-1999	99US-0149426P.			
PR	20-AUG-1999	99US-0149722P.			
PR	20-AUG-1999	99US-0149723P.			
PR	20-AUG-1999	99US-0149929P.			
PR	23-AUG-1999	99US-0149902P.			
PR	25-AUG-1999	99US-0150566P.			
PR	26-AUG-1999	99US-0150884P.			
PR	26-AUG-1999	99US-0151065P.			
PR	27-AUG-1999	99US-0151066P.			
PR	27-AUG-1999	99US-0151080P.			
PR	30-AUG-1999	99US-0151303P.			
PR	31-AUG-1999	99US-0151438P.			
PR	01-SEP-1999	99US-0151930P.			
PR	07-SEP-1999	99US-0152363P.			
PR	10-SEP-1999	99US-0153070P.			
PR	13-SEP-1999	99US-0153758P.			
PR					

```
QY 747 TAAAGAAACAATGAGCTTTCATGAGTCTTGATCAGATGTTCTTGATGATGCTCGGA 806
Db 113 -----AsnGlySerLeuSerValHisGlnHisHisSerAspGlnLeuPr 129
QY 807 GATAAGCTTTTGTGACAGCTACAGATTACCTCAGACGAGCTTCTTCAGCAGCAAGA 866
Db 129 oSerSerVal-----TyrAsnAs 135
QY 867 CATTTCTAATAAGCTGTGTT-----ACTCAAGGTTTCTC 899
Db 135 nAsnGlyAsnAsnGlyValGlyPheTyrAsnAsnTyrArgTyrGluThrSerGlyPheVa 155
QY 900 TCACTTATATTGGCTCAAAATACCTTCACCTCTGTTCAAGAAATACCTATCTCACTTCGC 959
Db 155 lSerSerValLeuArgSerArgTyrLeuLysProThrGlnLeuLeuAspGluValVa 175
QY 960 CGCATACTCGCTCGATTAT-----TCATCTCGAGGAACCGA 995
Db 175 lSerValArgLysAspLeuLysLeuGlyAsnLysLysMetLysAsnAspLysGlyGlnAs 195
QY 996 GTCAGAGCTGCTAGTTCAGCTTTTACTTCACGTTTTCAGATATATACTGAGTTTCTTGA 1055
Db 195 pPheHisAsnGlySerSer-----AspAsnIleThrGlu-----As 207
QY 1056 TGGTGATTCATACTCGAGCGGGTTTCGGATCTACATTTCAAAGGAGCATTTAGA 1115
Db 207 pAspLysSerGlnSerGlnGluLeu-----SerProSerGluArgGlnGluLeuGl 224
QY 1116 AGCAAGAAACCCATCTCTTGATCTCTTCAAAATGGTGGATGATCATGATGATGCT 1175
Db 224 nSerLysLysSerLysLeuLeuThrMetValAspGluValAspLysArgTyrAsnGlnTy 244
QY 1176 CTAAGATGAGATCATACGTTATATACGTTTTCAGCTTCATGCTCAACCGAGTTAGAT--CC 1232
Db 244 rHisHisGlnMetGluAlaLeuAlaSerSerPheGluMetValThrGlyLeuGlyAlaAl 264
QY 1233 ACAGTTACACACCCGGTTGGCTCCCAACCGTTTCCTTCTTATACAAAGACCTGAGAGA 1292
Db 264 aLysProTyrThrSerValAlaLeuAsnArgIleSerArgHisPheArgCysLeuArgAs 284
QY 1293 GAGATCTGCAAGAAGATA---ATCTCTATGGATCTCTATTTGAGAGAGAGCAAGACAA 1349
Db 284 pAlaIleLysGlnIleGlnValIleArgGlyLysLeuGlyGluArgGluThrSerAs 304
QY 1350 GACTCAAGAAACCTCTATGTTCCACCACCATGCTTCTTCAGCAGCTGAACCAAGAA 1409
Db 304 pGluGlnGlyGluArgIleProArgLeuArgTyrLeuAspGlnArgLeuArgGlnArg 324
QY 1410 C-----CATCAG-----ATTGGAGACCTCAACGAGTTTGC 1442
Db 324 gAlaLeuHisGlnGlnLeuGlyMetValArgProAlaTrpArgProGlnArgGlyLeuPr 344
QY 1443 TGAGAAATCTGTTTCGTTCTACGGAATGATGTTCCAAACTTCCTTCACCTTACCC 1502
Db 344 oGluAsnSerValSerIleLeuArgAlaTrpLeuPheGluHisPheLeuHisProTyrPr 364
QY 1503 GAAAGATCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAAGATCAGGTATC 1562
Db 364 oLysGluSerGlnLysIleMetLeuSerLysGlnThrGlyLeuSerLysAsnGlnValAl 384
QY 1563 AAACCTGGTTTAAATGCGGGTTAGCTATGAGCCGATGATAGAGATGATGATGC 1622
Db 384 aAsnTrpPheIleAsnAlaA-gValArgLeuTrpLysProMetIleGluGluMetTyLy 404
QY 1623 GGAA-----ATGAAACAGAGGAAGCTCAATAACAGTCAATTCACCCCAACGACC 1673
Db 404 sGluGluPheGlyGluSerAlaGluLeuLeuSerAsnSer-----AsnGlnAs 420
QY 1674 AACTCTTCGATGCAAAATCTGTTATGATGAGCCCAATGATGATGATGATGATGATGAT 1722
Db 420 pThrLysLysMetGlnGluThr-----SerGlnLeuLysHisGluAspSerSerSe 437
```

```
QY 1723 -----TAAGACAACAATTGTGTTTACCAACTTTGTGATAATTA 1760
Db 437 rSerGlnGlnGlnAsnGlnGlyAsnAsnAsnAsnAsnIleProTyrThrSerAsp-AlaG 457
QY 1761 GGCATTTGCTACTCTCTATGATTGCCAAAACCTTAACCATGTAGGACT 1807
Db 457 luGlnAsnLeuValPheAlaAspProLysProLysProAspArgAlaThrThr 472
RESULT 6
AAG14897
ID AAG14897 standard; protein; 524 AA.
XX
AC AAG14897;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 14932.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000BP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 03-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 23-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
```

```

Length:      524
Matches:     163
Conservative: 74
Mismatch:    170
Indels:      150
Gaps:        21

```



```

Db      263 rHisGlnMetGluAlaLeuAlaSerPheGluMetValThrGlyLeuGlyAlaA 283
QY      1233 ACAGTTATACACCGCGTTGCGCTCCAAACCGTTTCCTTATACAAAGAACCTTGAGAGA 1292
Db      283 alyProTyThrSerValAlaLeuAsnArgIleSerArgHisPheArgCysLeuArgAs 303
QY      1293 GAGATCTCCAGAGATA---ATCTTATGGAGTCTGTATTGGAGAGAGCAAGACAA 1349
Db      303 palalleysGluGlnIleGlnValleArgGlyLeuGlyGluArgGluThrSerAs 323
QY      1350 GACTCAAGAAACCTCTATGTTCCACCAGCATTCCTCTCAGCAGCTGAACAGAAAGAA 1409
Db      323 pGluGlnGlyGluArgIleProArgLeuArgTyLeuAspGlnArgLeuArgGlnArg 343
QY      1410 C-----CATCAG-----ATTGGAGACCTCAACGAGTTTGC 1442
Db      343 galAlaLeuHisGlnGlnLeuGlyMetValArgProAlaTrpArgProGlnArgGlyLeuPr 363
QY      1443 TGAGAAATCTGTTTCGGTTCTACGGAATTGGATGTTCCAAACTTCCTTCACCTTACCC 1502
Db      363 cGluAsnSerValSerIleLeuArgAlaTrpLeuPheGluHisPheLeuHisProTyPr 383
QY      1503 GAAAGATTGCGAGAAACATCTTCTAGCTATACGAGTGGCTTGACAAAGATCAGGTATC 1562
Db      383 olysGluSerGluLysIleMetLeuSerLysGlnThrGlyLeuSerLysAsnGlnValAl 403
QY      1563 AAACCTGTTTAAATGCGCGTTAGGCTATGAGCGCGATGATAGAGAGATGTATGC 1622
Db      403 asnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluGluMetTyLy 423
QY      1623 GGAA-----ATGACACAGAGGAAGTCAATACAGTGCACATTCAACCAACGCGACC 1673
Db      423 sGluGluPheGlyGluSerAlaGluLeuSerAsnSer-----AsnGlnAs 439
QY      1674 AACTCTTGATGCGAAATCTGTTATGATGAGCGCAACCAATGATATAA----- 1722
Db      439 pThrLysLysMetGlnGluThr-----SerGlnLeuLysHisGluAspSerSerSe 456
QY      1723 -----TAAGACACAAATTTGTTTACCAACTTTGTGATATAATTA 1760
Db      456 rSerGlnGlnAsnGlnGlyAsnAsnAsnAsnIleProTyThrSerAsp-Alag 476
QY      1761 GCCAATGCTACTCTATGATTTGCCCAAAACCTTAAACCATGTACCACT 1807
Db      476 luGlnAsnLeuValPheAlaAspProLysProAspArgAlaThrThr 491

RESULT 9
AAE02510
ID   AAE02510 standard; protein; 739 AA.
XX
AC   AAE02510;
DT
DT   10-AUG-2001 (first entry)
DE   Arabidopsis thaliana transcription factor G418.
XX
KW   Transcription factor; pesticidal; antimicrobial; gene therapy;
KW   pathogen tolerance; trichome structure; callose induction;
KW   phytoalexin induction; plant structure; plant development.
XX
OS   Arabidopsis thaliana.
XX
FH   Key
FT   Region
FT   500..560
FT   /note="Conserved domain"
XX
PN   WO200135726-A1.
XX
PD   25-MAY-2001.
XX
PF   14-NOV-2000; 2000WO-US031418.
XX

```

```

PR      17-NOV-1999; 99US-0166228P.
PR      17-APR-2000; 2000US-0197899P.
PR      22-AUG-2000; 2000US-0227439P.
XX
XX      (MEND-) MENDEL BIOTECHNOLOGY INC.
XX      (HEAR/) HEARD J.
XX      (RATC/) RATCLIFFE O.
XX      (CREE/) CREELMAN R.
XX      (JIAN/) JIANG C.
XX      (PINE/) PINEDA O.
XX      (REUB/) REUBER L.
XX      (ADAM/) ADAM L.
XX
PI      Heard J, Ratcliffe O, Creelman R, Jiang C, Pineda O, Reuber L;
PI      Adam L;
DR      WPI; 2001-335978/35.
DR      N-PSDB; AAD06479.
XX
XX      Nucleic acids encoding plant transcription factor polypeptides, useful
XX      for altering the pathogen resistance characteristics of plants, e.g.
XX      corn, potato and cotton plants.
XX
PS      Claim 4; Page 78-81; 134pp; English.
XX
XX      The present sequence is Arabidopsis thaliana transcription factor. The
XX      transcription factors are used to alter the structure and developmental
XX      characteristics of plants such as soybean, wheat, corn, potato, cotton,
XX      rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana,
XX      blackberry, blueberry, strawberry, raspberry, carrot, cantaloupe,
XX      cauliflower, coffee, cucumber, eggplant, grapes, mango, lettuce,
XX      honeydew, melon, onion, papaya, peas, peppers, pineapple, spinach,
XX      squash, sweet corn, tobacco, tomato, watermelon, roseaceous fruits and
XX      vegetable brassicas. The transcription factors are specifically useful
XX      for modifying traits associated with plant's pathogen tolerance such as
XX      alterations in cell wall composition, trichome number or structure,
XX      callose induction, phytoalexin induction, and alterations in the cell
XX      death response. Transgenic plants expressing these transcription factors
XX      are more tolerant to biotrophic or necrotrophic pathogens such as fungi,
XX      bacteria, molluscs, viruses, nematodes and parasitic higher plants. The
XX      transcription factors are also used in gene therapy
XX
SQ      Sequence 739 AA;
XX
Alignment Scores:
Pred. No.: 1,21e-34 Length: 739
Score: 424.00 Matches: 155
Percent Similarity: 41.02% Conservative: 94
Best Local Similarity: 25.54% Mismatches: 171
Query Match: 12.91% Indels: 187
DB: 4 Gaps: 24
XX
US-09-423-575-1 (1-1885) x AAE02510 (1-739)
QY      289 TTTTCATAGAACCCATCGACACAAAC----- 315
Db      17 PheHisAsnAsnSerMetSerGlnAspTyHisHisHisHisHisHisGlnHisGln 36
QY      316 -----AACAAACAACACTTTTATGTTCTCTGGAT-----AATGTCATGACT 357
Db      37 GlyGlyIlePheAsnPheSerAsnGlyPheAspArgSerAspSerProAsnLeuThrThr 56
QY      358 AACCAAAATCCT-----CTTCTCATGGAT----- 381
Db      57 GlnGlnLysGlnGluHisGlnArgValGluMetAspGluSerValalaGlyGly 76
QY      382 TTTTATACCTTCAAGAGAGATTCAACTTCATTCCTCAACAATGCTTCCTCCGAATACCATC 441
Db      77 ArgIleProValTyGluSerAlaGlyMetLeuSerGluMetPheAsnPhe----- 93
QY      442 AGATCAGATCCTCTCAAAATGGGTGGC-----TTTGATATTTTCAATCTATGCTG 492
Db      94 -----ProGlySerSerGlyGlyArgAspLeuAspLeuGlyGlnSerPheArg 110

```

QY	493	ACTAACAAATACCTTATCATCTTCCAGGCTCTATCGATGTTCAAGATAACCGCATGTT	552
Db	111	SerAsnArgGlnLeuGlu	123
QY	553	GAGTTCATG	561
Db	124	ProAlaMetAsnAlaThrAspSerAlaThrAlaAlaMetGlnLeuPheLeu	143
QY	562	-----GCTCCTCCTCATCCTCCCTCCA	585
Db	144	MetAsnProProGlnGlnProProSerProSerProSerThrThrSerProArgSer	163
QY	586	-----CTTCAT	612
Db	164	HisHisAsnSerSerThrLeuHisMetLeuLeuProSerProSerThrAsnThrThrHis	183
QY	613	TATGATCATTCCTCAACCAACATG	636
Db	184	HisGlnAsnTyrThrAsnHisMetSerMetHisGlnLeuProHisGlnHisGlnGln	203
QY	637	-----TGGGTTTGAACCAATAGTAGGTTTCAGGCAITTCAGGTAGTTGGT	687
Db	204	IleSerThrTrpGlnSerSerProAspHisHisHisHisAsn	219
QY	688	CCAAGTGAACCAATGATGTACATCTCGGTGAAGAAGATTCGCGTTCTTAATTCGAAT	747
Db	220	--SerGlnThrGluIleGlyThrValHisValGluAsn	234
QY	748	AAAGAAACAATGAGCTTTCAATGAGTCTTCATCAGATGTT	792
Db	235	HisGlyGlyGlnGlyLeuSerLeuSerLeuSerSerLeuGluAlaAlaAlaLysAla	254
QY	793	GATCAATGCTCGGAGATAAGTCTTTGTCAGCTACAAGATTAGCCTCA	840
Db	255	GluGluTyrArgAsnIleTyrTyrGlyAlaAsnSerSerAsnAlaSerProHisHisGln	274
QY	841	-----GAGCAAGCTTCTGCAGCAGCAAGACATTTCTAATAACGTTGTACT	888
Db	275	TyrAsnGlnPheLysThrLeuLeuAlaAsnSerSerGlnHisHisGlnValLeuAsn	294
QY	889	CAAGTTTCTCT	912
Db	295	GlnPheArgSerSerProAlaAlaSerSerSerMetAlaAlaValAsnIleLeuArg	314
QY	913	GGCTCAAAATACCTTCACTTGTTCAAGAAATATCTATCTCATTTCCGCGATCATCGCTC	972
Db	315	AsnSerArgTyrThrThrAlaAlaGlnGluLeuGluPheCysSer--ValGlyA	334
QY	973	GATTATTCAATCGAGGAACCGAGTCAAGAGCTGCTAGTTCAGCGCTTTACTCACGTTTT	1032
Db	334	rgGlyPheLeuLysLysAsnLysLeu-GlyAsnSerSerAsnProAsnThrCysGlyGly	353
QY	1033	GAGAATATACTGAGTTCCTTGATGGTCATTTCTAATACTCGGAGGCGGTTTCGGA	1089
Db	354	AspGly-----GlyLysSerSerProSerSerAlaGlyAlaAsnLys	367
QY	1090	-----TCTACATTTCAAGGAGAGACATTAGAAGCAAAACCCATCTC	1134
Db	368	GluHisProProLeuSerAlaSerAspArgIleGluHisGlnArgAlaGlyValLysLeu	387
QY	1135	TTGATCTTCTCAATGCTGATGATGATATAGTCATTCGCTAGATGAGATTCATACG	1194
Db	388	LeuThrMetLeuGluGluValAspArgArgTyrAsnHisTyrCysGluGlnMetGlnMet	407
QY	1195	GTTATATCAGCGTTC-----CATGCTGAACCGAGTTAGATCCACAGTTA	1239
Db	408	ValValAsnSerPheAspIleValMetClyHisGlyAlaAla-----LeuPro	423
QY	1240	CACACCGGTTTGGCTCCAAACCGTTTCCTTCTTATACAAACCTGAGAGAGAAATC	1299
Db	424	TyrThrAlaLeuAlaGlnLysAlaMetSerArgHisPheArgCysLeuLysAspAlaVal	443

Qy	1300	TGCAAGAGATA-----ATCTCT	1311
Db	444	AlaAlaGlnLeuLeuGlnSerCysGluLeuLeuGlyAspLysAspAlaAlaGlyIleSer	463
Qy	1318	ATGGGATCTGTATTGGACGAGGC-----AAAGACAAGACT	1353
Db	464	--SerSerGlyLeuThrLysGlyGluThrProArgLeuArgLeuGluGlnSerLeu	482
Qy	1354	CAGAACAACCTTATGTTCCACACAGATTCGCTTCTTCAGCAGCTGAACGAAAGAACCAT	1413
Db	483	ArgGlnGlnArgAlaPheHisGlnMetGlyMetMetGluGln-----	496
Qy	1414	CAGATTGGAGACCTCAACGAGGTTTGCTCGAGAAATCTGTTTCGGTTCTACGGAATTGG	1473
Db	497	GluAlaTrpArgProGlnArgGlyLeuProGluArgSerValAsnIleuAlaArgAlaTrp	516
Qy	1474	ATGTTCCAAACTCTCTTCACCCCTTACCCGAAAGATTCGGAGAAAATCTCTTAGCTATA	1533
Db	517	LeuPheGluHisPheLeuHisProTyrProSerAspAlaAspLysHisLeuLeuAlaArg	536
Qy	1534	CGAAGTCGGTTGACACAGACTCAGGTATCAAACTGGTTTATTAATCGCGGGTAGGCTA	1593
Db	537	GlnThrGlyLeuSerArgAsnGlnValSerAsnTrpPheIleAsnAlaArgValArgLeu	556
Qy	1594	TGGAAGCCGATGATAGACAGATGTATGCGGAAATGAACAAGAGAGCTCAATAACAGT	1653
Db	557	TrpLysProMetValGluGluMetTyrGlnGlnGluSerLysGluArgGluArgGluGlu	576
Qy	1654	CACATTCACCCCAAC	1668
Db	577	GluLeuGluGluAsn	581
RESULT 10			
ID	ADE37297	standard; protein; 739 AA.	
XX	AC	ADE37297;	
XX	AC		
XX	XX		
DT	29-JAN-2004	(first entry)	
XX	DE	Plant yield related protein from clone G418.	
XX	KW	transcription factor; tolerance; environmental condition;	
KW	KW	microbial disease; fungal disease; viral disease; pest infestation;	
KW	KW	herbicide sensitivity; heavy metal tolerance; heavy metal uptake;	
KW	KW	growth improvement; photocondition; nutrient uptake; hormone sensitivity;	
KW	KW	transgenic plant.	
XX	OS	Arabidopsis thaliana.	
XX	XX		
XX	FN	WO2003014327-A2.	
XX	XX		
PD	PD	20-FEB-2003.	
XX	XX		
PF	PF	09-AUG-2002; 2002WO-US026966.	
XX	XX		
PR	PR	09-AUG-2001; 2001US-0310847P.	
PR	PR	19-NOV-2001; 2001US-0336049P.	
PR	PR	11-DEC-2001; 2001US-0338692P.	
PR	PR	14-JUN-2002; 2002US-00171468.	
XX	XX		
PA	PA	(MEND-) MENDEL BIOTECHNOLOGY INC.	
XX	XX		
PI	PI	Reuber TL, Rischmann JL, Heard JE, Jiang C, Adam LJ, Dubell AN;	
PI	PI	Ratcliffe O, Pineda O, Yu GL, Broun PE;	
XX	XX		
XX	XX	WPI; 2003-256576/25.	
DR	DR	N-PSDB; ADE37296.	
XX	XX		
PT	PT	New stress-related transcription factor polynucleotides and polypeptides,	
PT	PT	useful for producing transgenic plants with e.g. improved tolerance to	
PT	PT	diseases or pests, decreased herbicide sensitivity, or improved nutrient	
PT	PT	uptake.	

XX Disclosure; SEQ ID NO 236; 470pp; English.

PS The invention relates to a number of cDNA sequence and their encoded

CC proteins which are especially transcription factor cDNAs and their

CC proteins. The isolated or recombinant polynucleotide is useful for

CC producing a modified plant with a modified trait, e.g. enhanced tolerance

CC to environmental conditions, improved tolerance to microbial, fungal or

CC viral diseases, improved tolerance to pest infestation, decreased

CC herbicide sensitivity, improved tolerance of heavy metals, or enhanced

CC ability to take up heavy metals, improved growth under poor

CC photoconditions, improved nutrient uptake, or reduced hormone

CC sensitivity. The transgenic plants are useful for growing a progeny plant

CC comprising the desired trait. The polynucleotides and polypeptides are

CC also useful in bioinformatic search methods. This sequence represents one

CC of the proteins of the invention.

XX

SQ Sequence 739 AA;

Alignment Scores:

Pred. No.: 1,21e-34 Length: 739

Score: 424.00 Matches: 155

Percent Similarity: 41.02% Conservative: 94

Best Local Similarity: 25.54% Mismatches: 171

Query Match: 12.91% Indels: 187

DB: 7 Gaps: 24

US-09-423-575-1 (1-1886) x ADE37297 (1-739)

QY	289	TTTCTAGAACCCAAATGGACAAAC	-----	315
Db	17	PheHisAsnAsnSerMetSerGlnAspTyrHisHisHisHisHisAsnGlnHisGln		36
QY	316	-----AACCAACAAACACTTTTAGTCTCTGGAT-----AATGTCATGACT		357
Db	37	GlyGlyIlePheAsnPheAsnGlyPheAspArgSerAspSerProAsnLeuThr		56
QY	358	AACCAAAATCCT-----CTTCTCATGGAT-----		381
Db	57	GlnGlnLysGlnGlnHisGlnArgValGluMetAspGluGluSerValAlaGlyGly		76
QY	382	TTTATACCTTCAGAGAGAGATCACTTCATCTCAACAATGCTTCCATGGAATACCATC		441
Db	77	ArgIleProValTyrGluSerAlaGlyMetLeuSerGluMetPheAsnPhe		93
QY	442	AGATCAGATCCTCTACAAATGGTGGC-----TTTGATATTTTCAATTTCTATGCTG		492
Db	94	-----ProGlySerSerGlyGlyArgAspLeuAspLeuGlyGlnSerPheArg		110
QY	493	ACTAACAAATATCTATCATCTTCCACGGTCTATCGATGTTCAAGATAACCGCAATGTT		552
Db	111	SerAsnArgGlnLeuLeuGlu-----GluGlnHisGlnAsnIle		123
QY	553	GAGTTTCATG-----CTTCAT-----		561
Db	124	ProAlaMetAsnAlaThrAspSerAlaThrAlaAlaAlaMetGlnLeuPheLeu		143
QY	562	-----GCTCTCTCTCTCATCTCTCTCTCA-----		585
Db	144	MetAsnProProProGlnGlnProProSerProSerSerThrThrSerProArgSer		163
QY	586	-----CTTCAT-----		612
Db	164	HisHisAsnSerSerThrLeuHisMetLeuLeuProSerProSerThrThrHis		183
QY	613	TATGATGATCTCTCAACAAACATG-----		636
Db	184	HisGlnAsnTyrThrAsnHisMetSerMetHisGlnLeuProHisGlnHisGlnGln		203
QY	637	-----TGGGGTTTGAAGCAAAATAGTCAGTTTCAGGCATTTTCAGGTGTAGTTGGT		687
Db	204	IleSerThrTrpGlnSerSerProAspHisHisHisHisHisAsn-----		219

QY	688	CCAAGTGAACCAATGATGCTCTACATTCGGTGAAGAGATTTCCCGTTCTTAATTTTCAAT	747
Db	220	---SerGlnThrGluIleGlyThrValHisValGluAsn-----SerGlyGly	234
QY	748	AAAAGAAACAATGAGCTTTTCATTGAGTCTTGATCAGATGTT-----TCT	792
Db	235	HisGlyGlyGlnGlyLeuSerLeuSerSerSerLeuGluAlaAlaAlaLysAla	254
QY	793	GATGAATGCTCGAGATAAGTCTTTGTGAGCTACAGATTAGCTCA-----	840
Db	255	GluGluTyrArgAsnIleTyrTyrGlyAlaAsnSerSerAsnAlaSerProHisGln	274
QY	841	-----GAGCAAGCTTCTTCAGCAGCAAAAGACATTTCTAATACCTGTGTACT	888
Db	275	TyrAsnGlnPheLysThrLeuLeuAlaAsnSerSerGlnHisHisGlnValLeuAsn	294
QY	889	CAAGTTTCTCT-----CAACTTATATT	912
Db	295	GlnPheArgSerSerProAlaAlaSerSerSerMetAlaAlaValAlaAsnIleLeuArg	314
QY	913	GGCTCAAAATACCTTCACCTCTGTTCAAGAAATACTATCTCATTTCCGCGCATACTCGCTC	972
Db	315	AsnSerArgTyrThrThrAlaAlaGlnGluLeuLeuGluPheCysSer--ValGlyA	334
QY	973	GATTATTCATCTCGAGAACCGAGTCAGAGCTGCTAGTTCAGCCTTTTACTTTCAGCTTTT	1032
Db	334	rgGlyPheLeuLysLysAsnLysLeu-GlyAsnSerSerAsnProAsnThrCysGlyGly	353
QY	1033	GAGAATATACCTGATTTCTTCATGCTTCTTAATACTCGAGCGGTTTCGA---1089	
Db	354	AspGly-----GlyGlySerSerProSerSerAlaGlyAlaAsnLys	367
QY	1090	-----TCTACATTTCAAAGGAGAGCATTAGAACAGAAAGAAACCCATCTC	1134
Db	368	GluHisProProLeuSerAlaSerAspArgIleGluHisGlnArgArgLysValLysLeu	387
QY	1135	TTGGATCTTTCAAATGGTGTGATCGATATAGTATCTTCATTCGCTAGTAGAGATTCATAG	1194
Db	388	LeuThrMetLeuGluGluValAspArgTyrAsnHisTyrCysGluGlnMetGlnMet	407
QY	1195	GTTATATCAGCGTTC-----CATGTCGACCGAGTGTAGATCCACAGTTA	1239
Db	408	ValValAsnSerPheAspIleValMetGlyHisGlyAlaAla-----LeuPro	423
QY	1240	CACACCGGTTTGGCTCCAAACCGTTCTTCCTTTATACAGAACCTGAGAGAGAAATC	1299
Db	424	TyrThrAlaLeuAlaGlnLysAlaMetSerArgHisPheArgCysLeuLysAspAlaVal	443
QY	1300	TGCAAGAAGATA-----ATCTCT	1317
Db	444	AlaAlaGlnLeuLysGlnSerCysGluLeuLeuGlyAspLysAspAlaAlaGlyIleSer	463
QY	1318	ATGGATCTGTATTCGAGAGAGGC-----AAACAAGACT	1353
Db	464	---SerSerGlyLeuThrLysGlyGluThrProArgLeuArgLeuGluGlnSerLeu	482
QY	1354	CAAGAAACCTCTATGTTCCACCGAGTTCCTTCTTCAGCAGCTGAAACGAAAGAACCAT	1413
Db	483	ArgGlnGlnArgAlaPheHisGlnMetGlyMetMetGluGln-----	496
QY	1414	CAGATTGGAGACCTCAACGAGTTTGGCTGAGAAATCTGTTTCGGTCTTACGGAATGG	1473
Db	497	GluAlaTrpArgProGlnArgGlyLeuProGluArgSerValAsnIleLeuArgAlaTrp	516
QY	1474	ATGTTCCAAAACCTTCTTCACCCCTTACCGAAAGATTCCGGAACATCTTCTAGCTATA	1533
Db	517	LeuPheGluHisPheLeuHisProTyrProSerAspAlaAspLysHisLeuLeuAlaArg	536
QY	1534	CGAAGTGGCTTCAACAAGAGTCAGGTATCAAACTGGTTTATAAATGCCGGGTAGGGTA	1593
Db	537	GlnThrGlyLeuSerArgAsnGlnValSerAsnTrpPheIleAsnAlaArgValArgLeu	556
QY	1594	TGGAACCGCATGATAGAGATGTATCGGAAATGAACAAGAGAACTCAATACAGT	1653

```
Db      557  TriPlysProMetValGluGluMetTyfGlnGlnGluSerLysGluArgGluGluGlu 576
QY      1654  CACATTCAACCAAC 1668
Db      577  GluLeuGluGluAsn 581
RESULT 11
AAG14140
ID  AAG14140 standard; protein; 459 AA.
AC      AC
XX      XX
XX      XX
DT      17-OCT-2000 (first entry)
DE      Arabidopsis thaliana protein fragment SEQ ID NO: 13890.
XX      XX
KW      Protein identification; signal transduction pathway; metabolic pathway;
KW      hybridisation assay; genetic mapping; gene expression control; promoter;
KW      termination sequence.
XX      XX
OS      Arabidopsis thaliana.
XX      XX
PN      EP1033405-A2.
PD      PD
XX      XX
XX      06-SEP-2000.
XX      XX
XX      25-FEB-2000; 2000EP-00301439.
XX      XX
XX      25-FEB-1999; 99US-0121825P.
XX      05-MAR-1999; 99US-0123180P.
XX      09-MAR-1999; 99US-0123548P.
XX      23-NAR-1999; 99US-0125788P.
XX      25-MAR-1999; 99US-0126264P.
XX      29-MAR-1999; 99US-0126785P.
XX      01-APR-1999; 99US-0127462P.
XX      06-APR-1999; 99US-0128234P.
XX      08-APR-1999; 99US-0128714P.
XX      16-APR-1999; 99US-0129845P.
XX      19-APR-1999; 99US-0130077P.
XX      21-APR-1999; 99US-0130449P.
XX      23-APR-1999; 99US-0130510P.
XX      23-APR-1999; 99US-0130891P.
XX      28-APR-1999; 99US-0134449P.
XX      30-APR-1999; 99US-0134048P.
XX      30-APR-1999; 99US-0134207P.
XX      04-MAY-1999; 99US-0134248P.
XX      05-MAY-1999; 99US-0134859P.
XX      06-MAY-1999; 99US-0134866P.
XX      07-MAY-1999; 99US-0134878P.
XX      11-MAY-1999; 99US-0134863P.
XX      14-MAY-1999; 99US-0134256P.
XX      14-MAY-1999; 99US-0134218P.
XX      14-MAY-1999; 99US-0134219P.
XX      14-MAY-1999; 99US-0134221P.
XX      18-MAY-1999; 99US-0134370P.
XX      18-MAY-1999; 99US-0134768P.
XX      19-MAY-1999; 99US-0134941P.
XX      20-MAY-1999; 99US-0135124P.
XX      21-MAY-1999; 99US-0135353P.
XX      24-MAY-1999; 99US-0135629P.
XX      25-MAY-1999; 99US-0136021P.
XX      27-MAY-1999; 99US-0136392P.
XX      28-MAY-1999; 99US-0136782P.
XX      01-JUN-1999; 99US-0137222P.
XX      03-JUN-1999; 99US-0137528P.
XX      04-JUN-1999; 99US-0137502P.
XX      07-JUN-1999; 99US-0137724P.
XX      08-JUN-1999; 99US-0138094P.
XX      10-JUN-1999; 99US-0138540P.
XX      10-JUN-1999; 99US-0138847P.
XX      14-JUN-1999; 99US-0139119P.
XX      16-JUN-1999; 99US-0139452P.
XX      16-JUN-1999; 99US-0139453P.
XX      17-JUN-1999; 99US-0139492P.
XX      18-JUN-1999; 99US-0139454P.
XX      18-JUN-1999; 99US-0139455P.
XX      18-JUN-1999; 99US-0139456P.
XX      18-JUN-1999; 99US-0139457P.
XX      18-JUN-1999; 99US-0139458P.
XX      18-JUN-1999; 99US-0139459P.
XX      18-JUN-1999; 99US-0139460P.
XX      18-JUN-1999; 99US-0139461P.
XX      18-JUN-1999; 99US-0139462P.
XX      18-JUN-1999; 99US-0139463P.
XX      18-JUN-1999; 99US-0139750P.
XX      18-JUN-1999; 99US-0139763P.
XX      21-JUN-1999; 99US-0139817P.
XX      22-JUN-1999; 99US-0139899P.
XX      23-JUN-1999; 99US-0140353P.
XX      23-JUN-1999; 99US-0140354P.
XX      24-JUN-1999; 99US-0140835P.
XX      28-JUN-1999; 99US-0140823P.
XX      29-JUN-1999; 99US-0140991P.
XX      30-JUN-1999; 99US-0141287P.
XX      01-JUL-1999; 99US-0141842P.
XX      02-JUL-1999; 99US-0142154P.
XX      02-JUL-1999; 99US-0142055P.
XX      06-JUL-1999; 99US-0142390P.
XX      08-JUL-1999; 99US-0142803P.
XX      09-JUL-1999; 99US-0142920P.
XX      12-JUL-1999; 99US-0142977P.
XX      13-JUL-1999; 99US-0143542P.
XX      14-JUL-1999; 99US-0143624P.
XX      15-JUL-1999; 99US-0144005P.
XX      16-JUL-1999; 99US-0144085P.
XX      19-JUL-1999; 99US-0144086P.
XX      19-JUL-1999; 99US-0144325P.
XX      19-JUL-1999; 99US-0144331P.
XX      19-JUL-1999; 99US-0144332P.
XX      19-JUL-1999; 99US-0144333P.
XX      19-JUL-1999; 99US-0144334P.
XX      19-JUL-1999; 99US-0144335P.
XX      20-JUL-1999; 99US-0144352P.
XX      20-JUL-1999; 99US-0144632P.
XX      20-JUL-1999; 99US-0144884P.
XX      21-JUL-1999; 99US-0144814P.
XX      21-JUL-1999; 99US-0145086P.
XX      21-JUL-1999; 99US-0145088P.
XX      22-JUL-1999; 99US-0145085P.
XX      22-JUL-1999; 99US-0145087P.
XX      22-JUL-1999; 99US-0145089P.
XX      23-JUL-1999; 99US-0145132P.
XX      23-JUL-1999; 99US-0145145P.
XX      23-JUL-1999; 99US-0145218P.
XX      23-JUL-1999; 99US-0145224P.
XX      26-JUL-1999; 99US-0145276P.
XX      27-JUL-1999; 99US-0145913P.
XX      27-JUL-1999; 99US-0145918P.
XX      27-JUL-1999; 99US-0145919P.
XX      28-JUL-1999; 99US-0145951P.
XX      02-AUG-1999; 99US-0146386P.
XX      02-AUG-1999; 99US-0146388P.
XX      02-AUG-1999; 99US-0146389P.
XX      03-AUG-1999; 99US-0147038P.
XX      04-AUG-1999; 99US-0147204P.
XX      05-AUG-1999; 99US-0147302P.
XX      05-AUG-1999; 99US-0147192P.
XX      05-AUG-1999; 99US-0147260P.
XX      06-AUG-1999; 99US-0147303P.
XX      06-AUG-1999; 99US-0147416P.
XX      09-AUG-1999; 99US-0147493P.
XX      09-AUG-1999; 99US-0147935P.
XX      10-AUG-1999; 99US-0148171P.
XX      11-AUG-1999; 99US-0148319P.
XX      12-AUG-1999; 99US-0148341P.
```


QY 1618 TATGGGAA 1626
Db ||| |||
329 TyrLysGlu 331

RESULT 12
AAG32446
ID AAG32446 standard; protein; 459 AA.
XX AC AAG32446;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 39142.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 03-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138847P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140895P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145244P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147935P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.


```

RESULT 13
AAG32445
ID AAG32445 standard; protein; 482 AA.
XX
AC AAG32445;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 39141.
XX
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX
25-FEB-1999; 99US-0121825P.
PR
05-MAR-1999; 99US-0121180P.
PR
09-MAR-1999; 99US-0123548P.
PR
23-MAR-1999; 99US-0125788P.
PR
25-MAR-1999; 99US-0126264P.
PR
29-MAR-1999; 99US-0126785P.
PR
01-APR-1999; 99US-0127462P.
PR
06-APR-1999; 99US-0128234P.
PR
08-APR-1999; 99US-0128714P.
PR
16-APR-1999; 99US-0128845P.
PR
19-APR-1999; 99US-0130077P.
PR
21-APR-1999; 99US-0130449P.
PR
23-APR-1999; 99US-0130510P.
PR
23-APR-1999; 99US-0130891P.
PR
28-APR-1999; 99US-0131449P.
PR
30-APR-1999; 99US-0132048P.
PR
04-MAY-1999; 99US-0132407P.
PR
04-MAY-1999; 99US-0132484P.
PR
05-MAY-1999; 99US-0132485P.
PR
06-MAY-1999; 99US-0132486P.
PR
06-MAY-1999; 99US-0132487P.
PR
07-MAY-1999; 99US-0132863P.
PR
11-MAY-1999; 99US-0134256P.
PR
14-MAY-1999; 99US-0134218P.
PR
14-MAY-1999; 99US-0134219P.
PR
14-MAY-1999; 99US-0134221P.
PR
14-MAY-1999; 99US-0134370P.
PR
18-MAY-1999; 99US-0134768P.
PR
19-MAY-1999; 99US-0134941P.
PR
20-MAY-1999; 99US-0135124P.
PR
21-MAY-1999; 99US-0135953P.
PR
24-MAY-1999; 99US-0135629P.
PR
25-MAY-1999; 99US-0136021P.
PR
27-MAY-1999; 99US-0136392P.
PR
28-MAY-1999; 99US-0136782P.
PR
01-JUN-1999; 99US-0137222P.
PR
03-JUN-1999; 99US-0137528P.
PR
04-JUN-1999; 99US-0137502P.
PR
07-JUN-1999; 99US-0137724P.
PR
08-JUN-1999; 99US-0138094P.
PR
10-JUN-1999; 99US-0138540P.
PR
10-JUN-1999; 99US-0138847P.
PR
14-JUN-1999; 99US-0139119P.
PR
16-JUN-1999; 99US-0139452P.
PR
16-JUN-1999; 99US-0139453P.
PR
17-JUN-1999; 99US-0139492P.
PR
18-JUN-1999; 99US-0139454P.
PR
18-JUN-1999; 99US-0139455P.
PR
18-JUN-1999; 99US-0139456P.
PR
18-JUN-1999; 99US-0139457P.
PR
18-JUN-1999; 99US-0139458P.
PR
18-JUN-1999; 99US-0139459P.
PR
18-JUN-1999; 99US-0139460P.
PR
18-JUN-1999; 99US-0139461P.
PR
18-JUN-1999; 99US-0139462P.
PR
18-JUN-1999; 99US-0139463P.
PR
18-JUN-1999; 99US-0139750P.
PR
18-JUN-1999; 99US-0139763P.
PR
21-JUN-1999; 99US-0139817P.
PR
22-JUN-1999; 99US-0139899P.
PR
23-JUN-1999; 99US-0140353P.
PR
23-JUN-1999; 99US-0140354P.
PR
24-JUN-1999; 99US-0140695P.
PR
28-JUN-1999; 99US-0140823P.
PR
29-JUN-1999; 99US-0140991P.
PR
30-JUN-1999; 99US-0141287P.
PR
01-JUL-1999; 99US-0141842P.
PR
01-JUL-1999; 99US-0142154P.
PR
02-JUL-1999; 99US-0142055P.
PR
06-JUL-1999; 99US-0142390P.
PR
08-JUL-1999; 99US-0142803P.
PR
09-JUL-1999; 99US-0142920P.
PR
12-JUL-1999; 99US-0142977P.
PR
13-JUL-1999; 99US-0143542P.
PR
14-JUL-1999; 99US-0143624P.
PR
15-JUL-1999; 99US-0144005P.
PR
16-JUL-1999; 99US-0144085P.
PR
16-JUL-1999; 99US-0144866P.
PR
16-JUL-1999; 99US-0144867P.
PR
19-JUL-1999; 99US-0144331P.
PR
19-JUL-1999; 99US-0144332P.
PR
19-JUL-1999; 99US-0144333P.
PR
19-JUL-1999; 99US-0144334P.
PR
19-JUL-1999; 99US-0144335P.
PR
20-JUL-1999; 99US-0144352P.
PR
20-JUL-1999; 99US-0144632P.
PR
20-JUL-1999; 99US-0144884P.
PR
21-JUL-1999; 99US-0144814P.
PR
21-JUL-1999; 99US-0145086P.
PR
21-JUL-1999; 99US-0145088P.
PR
22-JUL-1999; 99US-0145085P.
PR
22-JUL-1999; 99US-0145087P.
PR
22-JUL-1999; 99US-0145089P.
PR
22-JUL-1999; 99US-0145192P.
PR
23-JUL-1999; 99US-0145145P.
PR
23-JUL-1999; 99US-0145218P.
PR
23-JUL-1999; 99US-0145224P.
PR
26-JUL-1999; 99US-0145276P.
PR
27-JUL-1999; 99US-0145913P.
PR
27-JUL-1999; 99US-0145918P.
PR
27-JUL-1999; 99US-0145919P.
PR
28-JUL-1999; 99US-0145951P.
PR
02-AUG-1999; 99US-0146386P.
PR
02-AUG-1999; 99US-0146388P.
PR
02-AUG-1999; 99US-0146389P.
PR
03-AUG-1999; 99US-0147038P.
PR
04-AUG-1999; 99US-0147204P.
PR
05-AUG-1999; 99US-0147302P.
PR
05-AUG-1999; 99US-0147192P.
PR
05-AUG-1999; 99US-0147260P.
PR
06-AUG-1999; 99US-0147303P.
PR
06-AUG-1999; 99US-0147416P.
PR
09-AUG-1999; 99US-0147493P.
PR
09-AUG-1999; 99US-0147935P.
PR
10-AUG-1999; 99US-0148171P.
PR
11-AUG-1999; 99US-0148319P.
PR
12-AUG-1999; 99US-0148341P.
PR
13-AUG-1999; 99US-0148565P.
PR
13-AUG-1999; 99US-0148684P.
PR
16-AUG-1999; 99US-0149388P.
PR
17-AUG-1999; 99US-0149175P.
PR
18-AUG-1999; 99US-0149426P.
PR
20-AUG-1999; 99US-0149722P.

```

```

20-AUG-1999; 99US-0149723P.
20-AUG-1999; 99US-0149929P.
23-AUG-1999; 99US-0149902P.
23-AUG-1999; 99US-0149930P.
25-AUG-1999; 99US-0150566P.
26-AUG-1999; 99US-0150884P.
27-AUG-1999; 99US-0151065P.
27-AUG-1999; 99US-0151066P.
27-AUG-1999; 99US-0151067P.
30-AUG-1999; 99US-0151303P.
31-AUG-1999; 99US-0151438P.
01-SEP-1999; 99US-0151930P.
07-SEP-1999; 99US-0152363P.
10-SEP-1999; 99US-0153070P.
13-SEP-1999; 99US-0153758P.
15-SEP-1999; 99US-0154018P.
16-SEP-1999; 99US-0154039P.
20-SEP-1999; 99US-0154779P.
22-SEP-1999; 99US-0155139P.
23-SEP-1999; 99US-0155486P.
24-SEP-1999; 99US-0155659P.
28-SEP-1999; 99US-0156458P.
29-SEP-1999; 99US-0156596P.
04-OCT-1999; 99US-0157117P.
05-OCT-1999; 99US-0157753P.
06-OCT-1999; 99US-0157865P.
07-OCT-1999; 99US-0158029P.
08-OCT-1999; 99US-0158232P.
12-OCT-1999; 99US-0158369P.
13-OCT-1999; 99US-0159293P.
13-OCT-1999; 99US-0159294P.
13-OCT-1999; 99US-0159295P.
14-OCT-1999; 99US-0159329P.
14-OCT-1999; 99US-0159330P.
14-OCT-1999; 99US-0159331P.
14-OCT-1999; 99US-0159637P.
14-OCT-1999; 99US-0159638P.
18-OCT-1999; 99US-0159584P.
21-OCT-1999; 99US-0160741P.
21-OCT-1999; 99US-0160767P.
21-OCT-1999; 99US-0160768P.
21-OCT-1999; 99US-0160770P.
21-OCT-1999; 99US-0160814P.
21-OCT-1999; 99US-0160815P.
22-OCT-1999; 99US-0160980P.
22-OCT-1999; 99US-0160981P.
22-OCT-1999; 99US-0160989P.
25-OCT-1999; 99US-0161404P.
25-OCT-1999; 99US-0161405P.
25-OCT-1999; 99US-0161406P.
26-OCT-1999; 99US-0161359P.
26-OCT-1999; 99US-0161360P.
26-OCT-1999; 99US-0161361P.
28-OCT-1999; 99US-0161920P.
28-OCT-1999; 99US-0161922P.
28-OCT-1999; 99US-0161992P.
28-OCT-1999; 99US-0161993P.
29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 3,32e-34
Score: 419.00
Length: 482
Matches: 118
Percent Similarity: 47.1%
Conservative: 53
Best Local Similarity: 32.5%
Mismatch: 120
Query Match: 12.7%
Indels: 72
DB: 3
Gaps: 11

US-09-423-575-1 (1-1886) x AAG32445 (1-482)

QY 619 GATTCCTCAACACATGCGGGTTTGAACCAATAGTAGTTTCAGGCATTTTCAGGT 678
Db 37 AspSerThrAsnVal-----SerProSerLysGluLeuValLeuSerSer 53
QY 679 GTAGTTGCTCAAGTGAACCAATGATG-----TCTACATTCGGTGAAGAA 723

```



```

PR 23-AUG-1999; 99US-0149303P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158023P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159323P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 3, 32e-34
Length: 482
Scores: 419.00
Matches: 118
Percent Similarity: 47.11%
Conservative: 53
Best Local Similarity: 32.51%
Mismatch: 120
Query Match: 12.76%
Indels: 72
DB: 3
Gaps: 11

US-09-423-575-1 (1-1886) x AAG14139 (1-482)

QY 619 GATTCCTCAACACATGTCGGGTTTGAACCAATAGTAGTTCAGGCATTTTCAGGT 678
Db 37 AspSerThrAsnAsnVal-----SerProSerLysGluLeuGlnValLeuSerSer 53
QY 679 GTAGTTGGTCCCAAGTGAACCAATGATG-----TCTACATTCCGGTGAAGAA 723
Db 54 LeuGlyGlyValSerGlnMetValGluLeuAspSerGlySerTrpArgAspGlnGlu 73

```

```

QY 724 GATTTCCCGTTTCTAATTTTCGAATAAAGAAACAAT-----GAGCTTTTCATTGAGT 774
Db 74 Asp-----AsnAspArgAsnArgPheProValMetArgLeuGly 87
QY 775 CTTGCATCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTTTTGCGACGTACAAGATTA 834
Db 88 LeuSerSerGlnIleGlu----- 93
QY 835 GCCTCAGACCAAGCTTCTTCGACGACCAAGACATTTCTAATAAC-----GTT 882
Db 94 -----ThrSerArgGlyAsnAsnAsnAsnGluTrpAlaThrGln 106
QY 883 GTTACTCAAGGTTTCTCTCAACTATATTGGCTCAAAATACCTTCACTCTCTTCAAGAA 942
Db 107 ValValSerGlyPheThrArgThrIleHisAsnSerLysTyrLeuLysAlaAlaGlnGlu 126
QY 943 ATACTATCTCATTTCGCCCATCTCGCCATCTCGCTCGATATTATCATCTCGAGGACCGAGTCAGGA 1002
Db 127 LeuLeuAspGluThrValAsnValLys----- 135
QY 1003 GCTGCTAGTTCAGCCTTTACTTCACTTTTGAGAATAATACTGAGTTTCTTGTGATGTGAT 1062
Db 136 LysAlaLeuLysGlnPheGlnProGluGlyAspLysIleAsnGluValLysGluLysAsn 155
QY 1063 TCTAATACTCGGAGCGGGTTTCGGATCTACATTTCAAAGAGAGACATTAGAACAAAG 1122
Db 156 LeuGlnThrAsnThrAlaGluIleProGlnAla---GluArgGlnGluLeuGlnSerLys 174
QY 1123 AAAACCATCTCTCGATCTTCTCAATGTTGGTGATGATCGATATAGTATAGTCATTCGCTAGAT 1182
Db 175 LeuSerLysLeuLeuSerIleLeuAspGluValAspArgAsnTyrLysGlnTyrTrpHis 194
QY 1183 GAGATTTCATCGGTTTATACGCTTCCATCGTCAACCGAGTTAGAT---CCACAGTTA 1239
Db 195 GlnMetGlnIleValValSerSerPheAspValIleAlaGlyCysGlyAlaAlaLysPro 214
QY 1240 CACACCGGTTTCCCTCCAAACCGTTTCTTCTTATACAGAACCTCGAGAGAGATC 1299
Db 215 TyrThrAlaLeuAlaLeuGlnThrIleSerArgHisPheArgCysLeuArgAspAlaIle 234
QY 1300 TCGAAGAGATAATC-----TCT 1317
Db 235 SerGlyGlnIleLeuValIleArgLysSerLeuGlyGlyGluGlnAspGlySerAspGly 254
QY 1318 ATGGGATCTGTTATCGAGAGAGCGGAAACAGACAGACTCAAGAAACCTCTATGTTCCACCAG 1377
Db 255 ArgGlyValGlyIleSerArgLeuArgAsnValAspGlnGlnVal-----ArgGln 271
QY 1378 CATTCGCTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTTCGAGACCTCAACGAGGT 1437
Db 272 GlnArgAlaLeuGlnArgLeuGlyValMetGlnProHisThrTrpArgProGlnArgGly 291
QY 1438 TTGCTGTAGAAATCTGTTTCGGTTCACGGAATGCGATGTTCCAAATCTTCTTCCACCT 1497
Db 292 LeuProAspSerSerValLeuValLeuArgAlaTrpLeuPheGluHisPheLeuHisPro 311
QY 1498 TACCCGAAAGATTTCGAGAAACATCTTCTAGCTATACGAAGTGGCTTCACAAGAGTCCAG 1557
Db 312 TyrProLysAspSerAspLysIleMetLeuAlaArgGlnThrGlyLeuSerArgGlyGln 331
QY 1558 GTATCAAACTGGTTTATAAATCGCGGTTTAGGCTATCGAAGCCGATATGAGAAGATG 1617
Db 332 ValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetValGluGluMet 351
QY 1618 TATCGGGA 1626
Db 352 TyrLysGlu 354

```

RESULT 15

AAG21086

ID AAG21086 standard; protein; 491 AA.

XX

AC AAG21086;

PR	18-JUN-1999	99JUS-01397500P
PR	18-JUN-1999	99JUS-01396317P
PR	18-JUN-1999	99JUS-0139817P
PR	21-JUN-1999	99JUS-0139899P
PR	22-JUN-1999	99JUS-0140033P
PR	23-JUN-1999	99JUS-0140354P
PR	23-JUN-1999	99JUS-0140354P
PR	24-JUN-1999	99JUS-0140695P
PR	24-JUN-1999	99JUS-0140823P
PR	28-JUN-1999	99JUS-0140991P
PR	29-JUN-1999	99JUS-0141287P
PR	30-JUN-1999	99JUS-0141842P
PR	01-JUL-1999	99JUS-0142154P
PR	01-JUL-1999	99JUS-0142055P
PR	02-JUL-1999	99JUS-0142597P
PR	05-JUL-1999	99JUS-0142300P
PR	08-JUL-1999	99JUS-0142863P
PR	09-JUL-1999	99JUS-0143270P
PR	12-JUL-1999	99JUS-0143297P
PR	13-JUL-1999	99JUS-0143542P
PR	14-JUL-1999	99JUS-0143624P
PR	15-JUL-1999	99JUS-0144005P
PR	16-JUL-1999	99JUS-0144085P
PR	16-JUL-1999	99JUS-0144086P
PR	18-JUL-1999	99JUS-0144325P
PR	19-JUL-1999	99JUS-0144331P
PR	20-JUL-1999	99JUS-0144384P
PR	19-JUL-1999	99JUS-0144332P
PR	19-JUL-1999	99JUS-0144333P
PR	19-JUL-1999	99JUS-0144334P
PR	19-JUL-1999	99JUS-0144335P
PR	20-JUL-1999	99JUS-0144352P
PR	20-JUL-1999	99JUS-0144632P
PR	20-JUL-1999	99JUS-0144884P
PR	21-JUL-1999	99JUS-0144814P
PR	21-JUL-1999	99JUS-0145086P
PR	21-JUL-1999	99JUS-0145088P
PR	22-JUL-1999	99JUS-0145085P
PR	22-JUL-1999	99JUS-0145087P
PR	22-JUL-1999	99JUS-0145089P
PR	22-JUL-1999	99JUS-0145192P
PR	23-JUL-1999	99JUS-0145145P
PR	23-JUL-1999	99JUS-0145218P
PR	23-JUL-1999	99JUS-0145224P
PR	26-JUL-1999	99JUS-0145276P
PR	27-JUL-1999	99JUS-0145913P
PR	27-JUL-1999	99JUS-0145918P
PR	27-JUL-1999	99JUS-0145918P
PR	28-JUL-1999	99JUS-0145951P
PR	02-AUG-1999	99JUS-0145951P
PR	02-AUG-1999	99JUS-0146386P
PR	02-AUG-1999	99JUS-0146388P
PR	02-AUG-1999	99JUS-0146389P
PR	03-AUG-1999	99JUS-0147038P
PR	03-AUG-1999	99JUS-0147204P
PR	04-AUG-1999	99JUS-0147302P
PR	05-AUG-1999	99JUS-0147192P
PR	05-AUG-1999	99JUS-0147260P
PR	06-AUG-1999	99JUS-0147303P
PR	08-AUG-1999	99JUS-0147416P
PR	09-AUG-1999	99JUS-0147493P
PR	09-AUG-1999	99JUS-0147935P
PR	10-AUG-1999	99JUS-0148171P
PR	11-AUG-1999	99JUS-0148319P
PR	11-AUG-1999	99JUS-0148341P
PR	13-AUG-1999	99JUS-0148565P
PR	13-AUG-1999	99JUS-0148684P
PR	16-AUG-1999	99JUS-0149368P
PR	17-AUG-1999	99JUS-0149175P
PR	18-AUG-1999	99JUS-0149426P
PR	20-AUG-1999	99JUS-0149722P
PR	20-AUG-1999	99JUS-0149723P
PR	20-AUG-1999	99JUS-0149529P
PR	23-AUG-1999	99JUS-0149902P
PR	23-AUG-1999	99JUS-0149930P
PR	25-AUG-1999	99JUS-0150566P
PR	26-AUG-1999	99JUS-0150884P

Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 202, App
Sequence 2, Appli
Sequence 15, Appl
Sequence 14, Appl
Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 6, Appli
Sequence 2, Appli
Sequence 478, App
Sequence 485, App
Sequence 478, App
Sequence 485, App
Sequence 4, Appli
Sequence 6, Appli
Sequence 10, Appl
Sequence 9, Appli
Sequence 9, Appli
Sequence 14, Appl
Sequence 13, Appl
Sequence 4, Appli
Sequence 25, Appli
Sequence 9, Appli

13 106.5 3.2 971 3 US-08-918-190-2
14 106.5 3.2 971 3 US-09-234-232-2
15 106.5 3.2 971 5 PCT-US96-09927-2
16 104.5 3.2 1153 4 US-09-363-137-202
17 104 3.2 596 4 US-09-752-165-2
18 104 3.2 990 2 US-08-645-193B-15
19 102.5 3.1 1153 3 US-08-665-574C-14
20 102.5 3.1 1153 3 US-08-946-994-14
21 100.5 3.1 1147 3 US-08-470-260-5
22 100.5 3.1 1147 3 US-08-471-491-5
23 100.5 3.1 1147 3 US-08-466-662-5
24 100.5 3.1 1863 1 US-08-798-691-4
25 100.5 3.1 1863 3 US-08-825-487A-4
26 100.5 3.1 1863 3 US-09-074-476-6
27 100.5 3.1 3289 2 US-08-477-451-2
28 100 3.0 505 4 US-09-620-405B-478
29 100 3.0 505 4 US-09-620-405B-485
30 100 3.0 505 4 US-09-604-287A-478
31 100 3.0 505 4 US-09-604-287A-485
32 100 3.0 505 4 US-09-834-759-478
33 100 3.0 505 4 US-09-834-759-485
34 100 3.0 3079 5 PCT-US94-00198-4
35 99.5 3.0 1056 4 US-09-710-092-6
36 99.5 3.0 1270 4 US-09-710-092-10
37 99.5 3.0 1426 4 US-08-801-263A-9
38 99.5 3.0 2512 2 US-09-102-248-9
39 99.5 3.0 2512 4 US-09-367-764-9
40 99 3.0 1153 1 US-08-097-997A-14
41 98 3.0 3665 2 US-08-222-617A-13
42 98 3.0 3712 2 US-08-222-617A-4
43 98 3.0 3712 2 US-08-222-617A-25
44 98 3.0 3712 2 US-08-222-617A-25
45 97.5 3.0 1169 1 US-08-620-171A-9

ALIGNMENTS

RESULT 1
US-09-621-976-4543
; Sequence 4543, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Joebert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 4543
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-4543

Alignment Scores:	9.08e-08	Length:	142
Pred. No.:	151.00	Matches:	31
Score:	55.67%	Conservative:	23
Percent Similarity:	31.96%	Mismatches:	31
Best Local Similarity:	4.60%	Indels:	12
Query Match:	4	Gaps:	2
DB:			
US-09-423-575-1 (1-1886) x US-09-621-976-4543 (1-142)			
QY 1324 TCTATTGGAGAGGCAAGCAAGCAACTTATGTTCCACGACATTGC 1383			
Db 31 SerileMetSerArgAsnAsnAlaAspThrGlyArgValLeuAlaLeuProGluHis--- 49			
QY 1384 CTCTTCAGCAGCTGCAACGAGAACCATCAGATTGGAGACCTCAACGAGGTTTGCT 1443			

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: September 2, 2004, 12:02:38 ; Search time 24.5 seconds
(without alignments)
7948.292 Million cell updates/sec

Title: US-09-423-575-1
Perfect score: 3284
Sequence: 1 attaggtataaaattgtgc.....ttaaaaaaaaaaaaaaaaaa 1886

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DBV=xlh
-O=/cgn2_1/USPTO.spool/US09423575/runat_01092004_161107_18624/app.query.fasta_1.2055
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPTXT=0 -UNITS=5bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09423575 @CGN 1.1 26 @runat_01092004_161107_18624 -NCPU=6 -ICFU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCUS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151	4.6	142	4	US-09-621-976-4543
2	124.5	3.8	2954	4	Sequence 4543, Ap
3	115	3.5	1186	2	Sequence 1, Appli
4	115	3.5	1186	2	Sequence 8, Appli
5	115	3.5	1186	3	Sequence 8, Appli
6	114.5	3.5	676	4	US-09-323-433A-8
7	112.5	3.4	3969	3	Sequence 4318, Ap
8	110	3.3	944	4	US-09-134-001C-4318
9	109.5	3.3	1584	4	US-08-061-376-5
10	108	3.3	522	4	US-09-449-285A-2
11	107	3.3	512	2	US-09-457-040B-27
12	106.5	3.2	971	1	US-09-107-532A-3678
					Sequence 3678, Ap
					Sequence 5, Appli
					Sequence 2, Appli

Db 50 -----LysLysLysArgLysGlyAsn-----LeuPro 58
QY 1444 GAGAAATCTGTTTCGGTTCTACGGAATTCGATGTTCCAAACATCTCTTACCCG 1503
Db 59 AlaGluSerValLysIleLeuArgAspTrpMetTyrLysHisArgPheLysAlaTyrPro 78
QY 1504 ARAAGATTCGAGAAACATCTTAGCTATACGAAGTGGGTTGACAGAAGTCAGGTATCA 1563
Db 79 SerGluGluGlnMetLysGlnMetLysThrAsnLeuSerLeuLeuArgIleSer 98
QY 1564 AACTGGTTTAAATCGCGGGTTAGGCTATGGAAGCCGATGATAGAAGAG 1614
Db 99 AsnTyrPheIleAsnAlaArgArgIleLeuProAspMetLeuGlnGln 115
RESULT 2
US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; FILE OF INVENTION: Chromosome Congression
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; EARLIER FILING DATE: 1998-09-10
; EARLIER FILING DATE: US 60/058,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURE:
; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
US-09-150-867-1
Alignment Scores:
Pred. No.: 0.000415 Length: 2954
Score: 124.50 Matches: 103
Percent Similarity: 34.74% Conservative: 78
Best Local Similarity: 19.77% Mismatches: 205
Query Match: 3.79% Indels: 135
DB: 4 Gaps: 21
US-09-423-575-1 (1-1886) x US-09-150-867-1 (1-2954)
QY 142 CTTTCACGAGAAAGAGATAAAT-----AAGCATTATCTTCTTCTTCAAGT 195
Db 1871 LeuSerGluA:GAspAspLeuAsnGlnSerLysGlnHisLeuPheSerGluIleGluThr 1890
QY 196 CACATTTTGGAAATTTTCATGCTAAATTCCTTTGGAACGTTGTTGTTGCTGAATCTT 255
Db 1891 LeuSerLeuSerLeuLysGluLysGluPheAlaLeuGluGlnAla-----GluLys 1907

QY 256 CCCAAAGGTTCTATCAGAGAAGAGATAAAGTTTCTATAGAAAACCAATGGACAACAAC 315
Db 1908 AsplysAlaAspAlaAlaArgLysThrIleAspIle----- 1919
QY 316 AACAAACAACAACACTTTTAGTTCTCTGGTAATGTCATGACTAACCAAAATCCTCTTCTC 375
Db 1920 -----ThrGluLysIleSerAsnIleGluGluGlnLeuLeuGlnIleAlaThrAsnLeu 1937
QY 376 ATGGATTTTATACCTTCAAGAGAAGATTCAACTTCATTCTCAACAATGCTTCATCGAAT 435
Db 1938 LysGluThrLeuTyrGluArgGluSerLeuIleGlnCysLysGluGlnLeuAlaLeuAsn 1957
QY 436 -----ACCATCAGATCAGATCTCTTACAATGGTGGCTTGTGATTT 477
Db 1958 ThrGluHisLeuArgGluThrLeuLysSerLysAspLeuAlaLeuGlyLysMetGluGln 1977
QY 478 TTCAATTCTATGCTGACTAACAAATCTTATCATCTTCTCCACGG-----TCTATCGAT 531
Db 1978 GluArgAspGluAlaAlaAsnLysValIleAlaLeuThrGluLysMetSerSerLeuGlu 1997
QY 532 GTTCAAGATAACCGCAATGTTGAGTTTCATGGCTCTCTCTCTCATCTCTCTCCACTTCAT 591
Db 1998 GluGlnIleAsnGluAsnValThrThrLeuLys----- 2008
QY 592 CCTTGGATCATTTAAGACACTATGATGATCTCTCAACAACATGTGGGTTTGAAGCA 651
Db 2009 -----GluGly 2010
QY 652 AATAGTGGTTTCAGGCATTTTCAGGTGAGTGTGTCCAAGTGAACCAATGATGTCTACA 711
Db 2011 GluGlyGluLysGluThrPheTyr---LeuGlnArgProSerLysGlnGlnSerSer 2029
QY 712 TTCGGTGAAGAAGATTCCTCGGTTTCTTAATTCGAATAAAGAAACAATGAGCTTTCATG 771
Db 2030 GlnMetGluGlu-----LeuArgGluSerLeuLysThrLysAspLeuGlnLeu 2045
QY 772 AGTCTTTCGA---TCAGATGTTCTGTGATGATGCTCGAGATAGTCTTGTGACCTACA 828
Db 2046 GluGluAlaGluLysGluIleSerGluAlaThrAsnGluIleLysAsnLeuThrAlaLys 2065
QY 829 AGATTAGCCTCAGACCAAGCTTTCAGCAGACCAAGACATTTCTTAATAACCTTCTACT 888
Db 2066 IleSerSerLeuGluGluGluIleLeuGlnAsnAlaSerIleLeuAsnGluAlaValSer 2085
QY 889 CAAGGTTTCTCTCACTTATATTTGGCTCAAAATACCTTCACTCTGTTCAAGAAATACTA 948
Db 2086 Glu-----ArgGluAsnLeuArgHisSerLysGlnGlnLeuVal 2098
QY 949 TCTCATTTCCGCGCATCTCGCTCGATATTCATCTCGAGAACCGAGTCAGAGCTGCT 1008
Db 2099 SerGluLeuGluGlnLeuSerLeuThrLysSerArg----- 2111
QY 1009 AGTTCAGCTTTTACTTCAAGTATTAACCTGAGTTTCTTGTGTTGATGATCTAAT 1068
Db 2112 AspHisAlaPheAlaGlnSerLysArgGluLysAspGluAlaValAsnLysIleAlaSer 2131
QY 1069 AACTCGAGCGGGTTTCGGATCTACATTTCAAGAGAGCATTTAGAGCAAGAAACC 1128
Db 2132 LeuAlaGlu-----GluIleLysIleLeuThrLysGluMetAsp 2144
QY 1129 CATCTCTTGGATCTTCTCAATGTTGATGATGATATAGTATGCTGCTAGATGAGATT 1188
Db 2145 GluPheArgAspSerLysGluSerLeuGlnGlnSerSerHisLeuSerGluGluLeu 2164
QY 1189 CATACGGTTATACGCTTCCATGCTGCAACCGAGTTAGATCCACAGTTACACACCGG 1248
Db 2165 CysThrTyr-----LysThrGluLeuGln----- 2172
QY 1249 TTGCCCCCAACCGTTTCTTCTTATACAGAACCTCGAGAGAGAGATCTGCAAGAG 1308
Db 2173 -----MetLeuLysGlnGlnLysGluAspIleAsnAsnLys 2184
QY 1309 ATAATCTCTATGGGATCTGTATTGGAGAGAGCAAGCAAGAACTCAAGAAACCTCTATG 1368


```

Db      945 IleProSerAspGlnGlnAsnGluMetVal-----ArgGlu 956
QY      1111 TTAGAAGCAAGAAACCCATCTCTGGATCTTCTCAATGGTGATCGATATAGT 1170
Db      957 LeuAspGly-----HisValLeuLysCysValLys-----AspGlnAsnGlyAsn 971
QY      1171 CAP-----TGCCTAGAT-----GAGATTCATACGGTTATATCA 1203
Db      972 HisValValGlnLysCysLeuGluCysValGlnProGlnSerLeuGlnPheIleLeuAsp 991
QY      1204 GCGTTCCATGCTGCAACCCAGTAGATCCACAGTTACACACCCGGTTGGCCCTCAAAACC 1263
Db      992 AlaPheLysGlyGlnVal-----PheAlaLeuSerThr 1002
QY      1264 GTTCTCTCTTATACAAAGAACCTGAGAGAGAGAAATCTGCAAGAGATAATCTCTATGGGA 1323
Db      1003 HisProTyr-----GlyCysArgValIleGlnArgIleLeu----- 1014
QY      1324 TCTGTTATCGAGAGAGGCAAGAACTGCTAAGAAACCTCTATGTTCCACAGCATTCG 1383
Db      1015 -----GlnHisCys 1017
QY      1384 CTT-----CTTCAGCAGCTGAAGAAAGAACCATCAGATTGG 1422
Db      1018 LeuProAspGlnThrLeuProIleLeuGluLeuHisGlnHisThrGluGlnLeuVal 1037
QY      1423 AGACCTCAACGAGGTTGGCTGAGAAATCTGTTTCGGTTCTACGGAATTGGATGTCGAA 1482
Db      1038 GlnAspGlnTyrGly-----AsnTyrValIleGln 1047
QY      1483 AACTTCCTT---CACCTTACCGGAGATTCCGAGAAACATCTCTCTAGCTATACGAAGT 1539
Db      1048 HisValLeuGluHisGlyArgProGluAspLysSerLysIleValAlaGluIleArgGly 1067
QY      1540 GGC----- 1542
Db      1068 AsnValLeuValLeuSerGlnHisLysPheAlaSerAsnValValGluLysCysValThr 1087
QY      1542 ----- 1542
Db      1088 HisAlaSerArgThrGluArgAlaValLeuIleAspGluValCysThrMetAsnAspGly 1107
QY      1543 -----TTGCAAGAAAGTCAGGTATCAAACTGATTAATAAATGCG 1591
Db      1108 ProHisSerAlaLeuTyrThrMetMetLysAspGlnTyrAlaAsnTyrValVal----- 1125
QY      1582 CGGGTTAGGCTATGAAGCCGATGATAGAGAGATGTATGCGAAATGAACAAGAGGAAG 1641
Db      1126 -----GlnLysMetIleAsp-----ValAlaGluProGlyGlnArgLys 1138
QY      1642 CTCATAACAGTCACATCAACCCAGGACCACTCTTCGA 1683
Db      1139 IleValMetHisLysIleArgProHisIleAlaThrLeuArg 1152

```

RESULT 4

```

US-08-396-001-8
; Sequence 8, Application US/08396001
; Patent No. 5919618
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Claus, James
; APPLICANT: Cole, Francesca
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence in
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA

```

```

; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,001
; FILING DATE: 28-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6408A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-396-001-8

```

```

Alignment Scores:
Pred. No.: 0.00271 Length: 1186
Score: 115.00 Matches: 108
Percent Similarity: 31.77% Conservative: 68
Best Local Similarity: 19.49% Mismatches: 180
Query Match: 3.50% Indels: 198
DB: 2 Gaps: 25
US-09-423-575-1 (1-1886) x US-08-396-001-8 (1-1186)

```

```

QY      319 AACAAACACACITTTAGTTCTCTGGNTAATGTCATGACTAACCAAAATCCTCTCTCATG 378
Db      698 AsnSerAsnThrGlySerGly----- 704
QY      379 GATTTTATACCTTCAAGAGAGATTCACTTCAATCTCAACAATGCTTCCATGGAATACC 438
Db      705 -----SerArgArgAspSerLeuThrGlySerAspLeuTyrIysArgThr 720
QY      439 ATCAGATCAGATCCTCTACAAATGGTGGCTTGTATTTCAATCTATGCTGACTAAC 498
Db      721 -----SerSerSerLeuThrProIleGlyHisSerPheTyrAsnGlyLeu----- 735
QY      499 AAATACTTATCATCTTCTCCAGGCTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTC 558
Db      735 -----SerPheSerSerSerProGlyProValGlyMet----- 746
QY      559 ATGGCTCTCTCTCTCATCTCTCCACTTCTAT-----CCTTGGATCATTTA 606
Db      747 -----ProLeuProSerGlnGlyProGlyHisSerGlnThrProProSerLeuSer 764
QY      607 AGACATCATGATGATTCCTCAAAACAACTGGGGGTTTTGAAGCAAAATAGT----- 657
Db      765 SerHisGlySerSerSerSerLeuAsnLeuGlyLeuThrAsnGlySerGlyArgTyr 784
QY      658 -----GAGTTTCAGGCATTTTTCAGGTGTA 681
Db      785 IleSerAlaAlaProGlyAlaGluAlaLysTyrArgSerAlaSerAlaSerLeu 804
QY      682 GTTGTCCTCAAGTGAACCAATGATGCTACA-----TTCGGTGAAGACATTC 729
Db      805 PheSerProSerSerThrLeuPheSerSerSerArgLeuArgTyrGlyMetSerAspVal 824
QY      730 -----CCGTTTCTAATTTTCGAATAAAGAAACAATGAGCTTTCATTGAGT 774
Db      825 MetProSerGlyArgSerArgLeuLeuGluAspPheArgAsnAsnArgTyrProAsnLeu 844
QY      775 CTTGTCATCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTT-----TGTGCA 822

```

845 GlnLeuArgGluIleAlaGlyHisIleMetGluPheSerGlnAspGlnHisGlySerArg 864
 QY 823 GCTACAGATAGCTCTCAGCAAGCTTCTGCAGCAGCAAGCAATTTCTAATAAGTT 882
 Db 865 PheIleGlnLeuLysLeuGluArgAlaThrProAlaGluArgGlnLeuValPheAsnGlu 884
 QY 883 GTTACTCAAGGTTTCTCAACTT-----ATATTGGCTCAAAATACCTCACTCT 933
 Db 885 IleLeuGlnAlaAlaTyGlnLeuMetValAspValPheGlyAsnTyValIleGlnLys 904
 QY 934 GTTCAAGAAATACATCTCTTCCTCCGCGATCTCTCGATTTATTCATCTCGAGGACC 993
 Db 905 PhePheGluPheGlySerLeuGluGlnLysLeuAlaLeuAlaGluArgIleArgGlyHis 924
 QY 994 GAGTCAGGAGCTGTAGTTCAGCCTTTACTTCACGT---TTTGAGATATAACTGAGTTT 1050
 Db 925 ValLeuSerLeuAlaLeuGlnMetTyGlyCysArgValIleGlnLysAlaLeuGluPhe 944
 QY 1051 CTTGATGGTGTATCTTAATAACTCGGAGCGGGTTTCGGATCTACATTTCAAGAGAGACA 1110
 Db 945 IleProSerAspGlnGluAsnGluMetVal-----ArgGlu 956
 QY 1111 TTGAAGCAAGAAACCATCTCTTGGATCTTCTCAATGGTGGATGATGATATAGT 1170
 Db 957 LeuAspGly-----HisValLeuLysCysValLys-----AspGlnAsnGlyAsn 971
 QY 1171 CAT-----TGCGTAGAT-----GAGATTCATACGGTTTATATCA 1203
 Db 972 HisValValGlnLysCysIleGluCysValGlnProGlnSerLeuGlnPheIleIleAsp 991
 QY 1204 GGGTTCCATCTGCAACCGAGTAGATCCACAGTTACACCGCGTTTCCTCCCAACC 1263
 Db 992 AlaPheLysGlyGlnVal-----PheAlaLeuSerThr 1002
 QY 1264 GTTCTCTCTTATACAGAACCTGAGAGAGAGAACTCGAAGAGATATCTCTATGGGA 1323
 Db 1003 HisProTyT-----GlyCysArgValIleGlnArgIleLeu----- 1014
 QY 1324 TCTGATTCGAGAGAGGCAAGACAGACTCAAGAAACCTCTATGTTCCACAGATGTC 1383
 Db 1015 -----GluHisCys 1017
 QY 1384 CTT-----CTTCAGAGCTGAAACGAAAGAACCAATCAGATTTGG 1422
 Db 1018 LeuProAspGlnThrLeuProIleLeuGluLeuHisGlnHisThrGluGlnLeuVal 1037
 QY 1423 AGACCTCAACGAGTTTGGCTCAGAAATCTGTTCCGGTTCTACGGAATGGATGTTCCAA 1482
 Db 1038 GlnAspGlnTyGly-----AsnTyValIleGln 1047
 QY 1483 AACTTCCTT---CACCTTACCGAAGATTGGAGAACATCTTCTAGCTATACGAAT 1539
 Db 1048 HisValLeuGluHisGlyArgProGluAspLysSerLysIleValAlaGluIleArgGly 1067
 QY 1540 GGC----- 1542
 Db 1068 AsnValLeuValLeuSerGlnHisLysPheAlaSerAsnValValGluLysCysValThr 1087
 QY 1542 ----- 1542
 Db 1088 HisAlaSerArgThrGluArgAlaValLeuIleAspGluValCysThrMetAsnAspGly 1107
 QY 1543 -----TTGACAAGAGTCAGGTATCAAACTGGTTTATAATGGC 1581
 Db 1108 ProHisSerAlaLeuTyThrMetMetLysAspGlnTyAlaAsnTyValVal----- 1125
 QY 1582 CGGGTTAGGCTATGGAAGCGCATGATAGAGAGATGATGCGGAATGAACAGAGGAAG 1641
 Db 1126 -----GlnLysMetIleAsp-----ValAlaGluProGlyGlnArgLys 1138
 QY 1642 CTCAATAACAGTCACATTCACCAACCGACCACTCTCGA 1683
 Db 1139 IleValMetHisLysIleArgProHisIleAlaThrLeuArg 1152

RESULT 5

US-09-323-433A-8
 ; Sequence 8, Application US/09323433A
 ; Patent No. 6218512
 ; GENERAL INFORMATION:
 ; APPLICANT: Guarente, Leonard P.
 ; APPLICANT: Austriaco Jr., Nicanor
 ; APPLICANT: Claus, James J.
 ; APPLICANT: Cole, Francesca
 ; APPLICANT: Kennedy, Brian
 ; TITLE OF INVENTION: GENES DETERMINING CELLULAR SENESENCE IN
 ; TITLE OF INVENTION: YEAST
 ; FILE REFERENCE: 0050.1491-003
 ; CURRENT APPLICATION NUMBER: US/09/323,433A
 ; CURRENT FILING DATE: 1999-06-01
 ; PRIOR APPLICATION NUMBER: US 08/396,001
 ; PRIOR FILING DATE: 1995-02-28
 ; PRIOR APPLICATION NUMBER: PCT/US94/09351
 ; PRIOR FILING DATE: 1994-08-15
 ; PRIOR APPLICATION NUMBER: US 08/107,408
 ; PRIOR FILING DATE: 1993-08-16
 ; NUMBER OF SEQ ID NOS: 48
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 8
 ; LENGTH: 1186
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-323-433A-8

Alignment Scores:
 Pred. No.: 0.00271 Length: 1186
 Score: 115.00 Matches: 108
 Percent Similarity: 31.77% Conservative: 68
 Best Local Similarity: 19.49% Mismatches: 180
 Query Match: 3.50% Indels: 198
 DB: 3 Gaps: 25

US-09-423-575-1 (1-1886) x US-09-323-433A-8 (1-1186)

QY 319 AACAAACACTTTTATGTTCTCTGGATAATGTCATGACTAACCAAAATCCTCTCTCATG 378
 Db 698 AsnSerAsnThrGlySerGly----- 704
 QY 379 GATTTTATACCTTCAAGAGAAGATTCAACTTCATTCTCAACAATGCTTCCATGGAATACC 438
 Db 705 -----SerArgArgAspSerLeuThrGlySerSerAspLeuTyLysArgThr 720
 QY 439 ATCAGATCAGATCCTCTACAAATGGGTGGGCTTTGATATTTTCAATTTCTATCTGACTAAC 498
 Db 721 -----SerSerSerLeuThrProIleGlyHisSerPheTyArgGlyLeu----- 735
 QY 499 AATACATTATCATCTTCTCCACGGTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTC 558
 Db 736 ---SerPheSerSerProGlyProValGlyMet----- 746
 QY 559 ATGGCTCTCTCTCTCATCTCTCTCCACTTCAT-----CCTTTGGATCATTTA 606
 Db 747 -----ProLeuProSerGlnGlyProGlyHisSerGlnThrProProSerLeuSer 764
 QY 607 AGACACTATGATGTTCTCTCAACAACATGTTGGGGTTTGAACCAATAGT----- 657
 Db 765 SerHisGlySerSerSerSerLeuAsnLeuGlyLeuThrAsnGlySerGlyArgTyTyr 784
 QY 658 -----CAGTTTCAGGCATTTTCAGGTGTA 681
 Db 785 IleSerAlaAlaProGlyAlaGluAlaLysTyArgSerAlaSerAlaSerSerLeu 804
 QY 682 GTTGGTCCAAGTGAACCAATGATCTCTACA-----TTCGGTGAAGAAGATTTTC 729
 Db 805 PheSerProSerSerThrLeuPheSerSerSerArgLeuArgTyGlyMetSerAspVal 824
 QY 730 -----CCGTTTCTAATTCGAATAAAGAAACAATGAGCTTTCATTGAGT 774


```
QY 772 AGTCTGCATCAGATGTT-----TCTGATGAATGCTCGGAG 807
Db 420 ThrSerSerAspAspIleLeuLysSerMetLeuAspGlnSerLysAspLysGluSerLeu 439
QY 808 ATAAGTCTTTGTGAGTGTACAGATTAGCTTCAGAGCAAGCTTCTTCAGCAGCAAGAGAC 867
Db 440 IleLysGlnLeuThrThrArgLeuGlyAsnAspGluAlaAspArgIleAlaLysLys 459
QY 868 ATTCTTAATACGTTGTACTCAAGGTTCTCTCAACTATATTTGGCTCAAAATACCTT 927
Db 460 Leu-----LeuSerGlnAsnLeuSerAsnSerGlnIle--- 470
QY 928 CACTCTGTTCAAGAAATACTATCTCATTTCCCGCATACTGCTCGATTATTCTATCTCGA 987
Db 471 -----ValGluGlnLeuLysArgHisPhe-----AsnSerGln 481
QY 988 GGAACCGAGTCAGAGCTGTAGTTTCAGCCTTTTACTTCACGTTTGTGAGATATTAACGTAG 1047
Db 482 GlyThrAlaThrAla-----AspAspIleLeuAsn 491
QY 1048 TTTCTGTGATGTTCTTAATAACTCGAGCGGGTTTCGGATCTACATTTCAAGAGAGA 1107
Db 492 GlyValIleAsnAspAlaLysAspLysArgGlnAlaIleGluThrIleLeuGlnThrArg 511
QY 1108 GCATTAGAGCAAGAAACCCATCTCTGTGATCTTCTTCAATGGTGGATGATCGATAT 1167
Db 512 IleAsnLysAspLysAlaLysIleIleAlaAspValIleAlaArgValGlnLysAspLys 531
QY 1168 AGTCATTCGTAGTACAGATTCATACGTTTATATCAGCTTCCATGCTGCAACCGAGTTA 1227
Db 532 SerAspIleMetAspLeuIleHisLeuAlaIleGluGly-----LysAlaAsnAspLeuLeu 550
QY 1228 GATCCACAGTTACACACCCGGTTTCCCTCCAAACCGTTCTCTTCTTATACAGAACCTG 1287
Db 551 AspIleGluLysArgAlaLysGlnAlaLysLysAspLeuGluTyIleLeuAspProIle 570
QY 1288 AGAGAGAGAACTGCAAGAGATATCTCTATGGATCTGTATTGGAGAGA---GGCAAA 1344
Db 571 LysAsnArg-----ProSerLeuLeuAspArgIleAsnLys 582
QY 1345 GACAGAGACTCAAGAAACCTGTATGTTCCACAGCATTCCTCTTTCAGCAGCTGAAACGA 1404
Db 583 GlyValGlyAspSerAsnSerIlePheAspArgProSerLeuLeuAspLysLeuHisSer 602
QY 1405 AAGAACCATCAGATTTGGAGACCTCAACGAGGTTTGCCTGAGAAATCTGTTTCGGTT 1461
Db 603 ArgGlySerIleLeuAspLysLeuAspHisSerAlaProGluAsnGlyLeuSerLeu 621
```

RESULT 7

```
US-08-061-376-5
; Sequence 5, Application US/08061376
; Patent No. 6175000
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Djabali, Malek
; APPLICANT: Salleri, Lucia
; APPLICANT: Parry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,376
; FILING DATE: 13-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9387
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)546-4737
; TELEFAX: (619)546-9392
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3969 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-061-376-5
```

Alignment Scores:

```
Pred. No.: 0.0102 Length: 3969
Score: 112.50 Matches: 130
Percent Similarity: 31.50% Conservative: 76
Best Local Similarity: 19.88% Mismatches: 215
Query Match: 3.43% Indels: 235
DB: 3 Gaps: 26
```

US-09-423-575-1 (1-1886) x US-08-061-376-5 (1-3969)

```
QY 39 TCTGAATCTTTAGTGGCAGATGATGAAGATTATGAATTTCTTCATGAAATATTGTA 98
Db 2385 ThrGlnSerAlaAsnSerSerProAspGluAspThrGluVal----- 2398
QY 99 AGAAAAAACATAGAGAAGCTGCGGAATGAAGTAACTGTTCTTTCACGGAGAAAGAA 158
Db 2399 -----LysThrLeuLysLeuSerGlyMetSer----- 2407
QY 159 GATAAATAGCATATCTCTCTTCAGTTTAAACACACATTTTGGAAATTTTGATGTA 218
Db 2408 -----AsnArgSerSerIleIleAsnGluHisMetGly----- 2418
QY 219 AAAATTCTCTTGGAACTGTTGTTCTCTGAAATCTTCCAAAGGTTCTCATGAGAAG 278
Db 2419 -----SerSerSerArgAspArgArgGlnLys 2427
QY 279 AAGGATAAAGTTTCATAGAAACCCATGGACAAACAAACAAACAAACACACTTTAGTTC 338
Db 2428 GlyLysLysSerCys-LysGluThrPhenylsGluLysHisSerSerLysSerPheLeuGl 2447
QY 339 TCTGGATAATGTCATGACT-----AACCAAAATCTCTTCTCATGATTTTAT 386
Db 2447 uProGlyGlnValThrThrGlyGluGluGlnLysProGluPheMetAspGluVal 2467
QY 387 ACCTTCAAGAGAAAGATTCAACTTCATTCACAAATGCTTCCATGGAATACCATCAATC 446
Db 2467 lLeuThrProGlu-----TyrMetGlyGlnArgProCysAsnAsnValSerSe 2483
QY 447 AGAT-----CCTCTACA 458
Db 2483 AspLysIleGlyAspLysGlyLeuSerMetProGlyValProLysAlaProMetGlu 2503
QY 459 AATCGGTGGC----- 468
Db 2503 nValGluGlySerAlaLysGluLeuGlnAlaProArgLysArgThrValThrLe 2523
QY 469 -----TTTGATATTTCAT---TCTATGCTGACTAACAAATATTATCATCTCTCC 518
Db 2523 uThrProLeuLysMetGluAsnGluSerGlnSerLysAsnAlaLeuLysGluSerSerPr 2543
QY 519 ACGGTCT---ATCGATGTTCAAGATAACCGCAATGTTGAGTTTCATGGCTCCTCCTCTCA 575
Db 2543 AlaSerProLeuGlnIleGluSerThrSerProThrGluProIleSerAlaSerGluAs 2563
```

QY	576	TCCT-----CCTCCACTTCATCCCTTTGGATCATTTAAGACACACTATGATGATTCCTC	626
Db	2563	nProGlyAspGlyProValAlaGlnProSerProAsnAsnThrSerCysGlnAspSerG1	2583
QY	627	AAACAACATGTGG-----	639
Db	2583	nSerAsnAsnTyrGlnAsnLeuProValGlnAspArgAsnLeuMetLeuProAspGlyPr	2603
QY	640	-----GGTTTGGAA-----GCAATAGTGA	659
Db	2603	oLysProGlnGluAspGlySerPheLysArgGlyTyrProArgArgSerAlaAlaGalaA	2623
QY	660	GTTTCAGGCATTTTCAGGTGTAGTTGGTCCAAGTGAACCAATG-----ANGTCTACATT	713
Db	2623	gSerAsnMetPhePheGlyLeuThr-----ProLeuTyrGlyValArgSerTyr	2639
QY	714	CGGTGAAGAAGATTTCCGTTTCTTAATTTCCGAAT-----AAAAGAAA	755
Db	2639	rGlyGluGluAspIleProPheTyrSerSerThrGlyLysLysArgGlyLysArgSe	2659
QY	756	CAATGAGCTTTTCATGTAGCTCTGCATCAGATGTTTCTCATGATCAATGCTCGAG-----	807
Db	2659	rAlaGluGlyGlnValAspGlyAlaAspAspLeuSerThrSerAspGluAspAspLeuTyr	2679
QY	808	-----ATAAGTCTTTGTGCAGCTACAAGATTAGCTCA-----	840
Db	2679	rTyrTyrAsnPheThrArgThrValIleSerSerGlyGlyGluArgLeuAlaSerHi	2699
QY	841	-----GAGCAAGCTTCTTGACGACGCAAAACACATTTCTTAATAACGTTGTTAC	887
Db	2699	sAsnLeuPheArgGluGluGlnCysAspLeuProLysIleSer-----	2714
QY	888	TCAAGGTTTCTCTCAACTTATATTGGCTCAAAATACCTTCACCTCTGTTCAGAAGAACTACT	947
Db	2714	-----	2714
QY	948	ATCTCATTTGCCGCATCTCGCTCGATTATTCATCTCGAGAACCGAGTCAGGAGCTGC	1007
Db	2715	-----GlnLeuAspGlyValAspGlyThrGluSerAspThrSe	2728
QY	1008	TAGTTCAGCCTTTACTTCACGTTTGGAGATATACTGAGTTTCTTCATGGTGATTCATA	1067
Db	2728	rValThrAlaThrThrArgLysSerSerGlnIleProLys---ArgAsnGlyLysGluAs	2747
QY	1068	TAACTCGGAGGC-----GGGTTTCGGATCTACATTTC	1099
Db	2747	nGlyThrGluAsnLeuLysIleAspArgProGluAspAlaGlyGlyLysGluHisValTh	2767
QY	1100	AAAGGAGAGANTTAGAGCAAGAAACCCATCTCTTGGATCTCTTCAATGGTGGATG	1159
Db	2767	rLysSerSerValGlyHisLysAsnGluProLys-MetAspAsnCysHisSerValSerA	2787
QY	1160	ATCGATATAGTCATTGCGTAGATTACATACGGTTATATACAGCGTTC-----	1209
Db	2787	rgValLysThrGlnGlyLysSerLeuGluAlaGlnLeuSerSerLeuGluSerSera	2807
QY	1210	-----CATGCTGCACCGAGTTAGATCCACAGTTTACACACCGGTTTGGCCTCCAA-	1260
Db	2807	rgArgValHisThrSerThrProSerAspLysAsnLeuLeuAspThrTyrAsnThrGluL	2827
QY	1260	-----	1260
Db	2827	euleuLysSerAspSerAspAsnAsnSerAspAspCysGlyAsnIleLeuProSera	2847
QY	1261	--ACCGTTTCTCTTTATACAAGAACCTG-----AGAG	1291
Db	2847	spIleMetAspPheValLeuLysAsnThrProSerMetGlnAlaLeuGlyLysSerProG	2867
QY	1292	AGAGAACTTCGAAGAAGATAATCTCTATGGATCTGTATTGGAGAGCGGCAAGACAGA	1351
Db	2867	luSerSerSerGluLeuLeuAsnLeuGlyGluGlyLeu-----GlyLeuAspSera	2885

QY	1352	CTCAAGAAACCTCTATGTTTCCACAGCATTCCTTCTAGCAGCTGAA-ACGAAAGAAC	1411
Db	2885	snArgGluLysAspMetGlyLeuPheGluValPheSerGlnGlnLeuProThrThrGluP	2905
QY	1411	CATCAGATTTCGGAGACCTCAACGAGGTTTCCTGAGAAATCTCTTCGGT-----	1460
Db	2905	roValAspSerSerValSerSerIleSerAlaGluGlnPheGluLeuProLeuG	2925
QY	1461	-----TCTACGGAATGGAGTT-----CCAAACTTCCTTCACCTTAC	1500
Db	2925	luLeuProSerAspLeuSerValLeuThrThrArgSerProThrValProSerGlnAsnP	2945
QY	1501	CCGAAAGATTTCGGAGAAACATCTTCTAGCTATACGAAG	1538
Db	2945	roSerArgLeuAlaValIleSerAspSerGlyGluLys	2957
RESULT 8			
US-09-449-285A-2			
; Sequence 2, Application US/09449285A			
; Patent No. 6313280			
; GENERAL INFORMATION:			
; APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnology			
; TITLE OF INVENTION: SMAD-INTERACTING POLYPEPTIDES AND THEIR USE			
; FILE REFERENCES: 2676-4232US			
; CURRENT APPLICATION NUMBER: US/09/449,285A			
; CURRENT FILING DATE: 1999-11-24			
; PRIOR APPLICATION NUMBER: PCT/EP98/03193			
; PRIOR FILING DATE: 1998-05-28			
; PRIOR APPLICATION NUMBER: 97201645.5			
; PRIOR FILING DATE: 1997-06-02			
; NUMBER OF SEQ ID NOS: 27			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 2			
; LENGTH: 944			
; TYPE: PRT			
; ORGANISM: Mus musculus			
US-09-449-285A-2			
Alignment Scores:			
Pred. No.:	0.00841	Length:	944
Score:	110.00	Matches:	106
Percent Similarity:	32.19%	Conservatives:	73
Best Local Similarity:	19.06%	Mismatches:	178
Query Match:	3.35%	Indels:	199
DB:	4	Gaps:	28
US-09-423-575-1 (1-1886) x US-09-449-285A-2 (1-944)			
QY	195	ACACATTTTGGAAATTTTCATGTAAATTCCTTTGGACCTTGTGTCTGAAATCT	254
Db	49	SerHisSerGlySerTyrSerSerHisile-SerSerLysLysCysile-----	64
QY	255	TCCCAAAGTTCATC-----AGAAAGAAGAGATAAAGTTTCATAGAAA	299
Db	65	-----GlyLeuIleSerValAsnGlyArgMetArgAsnIleLysThrGlySerSe	82
QY	300	CCCA-----ATGGACACACACACACAACTTTAGTCTCTCGGATAATGTCAT	353
Db	82	rProAsnSerValSerSerProThrAsnSerAlaIleThrGlnLeuArgAsnLysLe	102
QY	354	GACTAACCAAAATCTCTTCTCATGGATTTTATACCTTCAAGAGAAGATTCAACTTCAT	413
Db	102	uGluAsnGlyLysProLeu-----SerMe	110
QY	414	CTCAACAATGCTTCATGAATACCATCAGATCAGATCCTCTACAAATGGGTGGCTTGA	473
Db	110	tSerGluGlnThrGlyLeuLeuLysIleLysThrGluProLeuAspPheAsnAspTyrLy	130
QY	474	TATTTTCAATTCATGCTGACTTAACAATACTTATCATCTTCTCCACCGTCTATCGATGT	533
Db	130	valLeu-----MetAlaThrHisGlyPheSerGlySerPro-----	143
QY	534	TCAAGTAACCCCAATGTTGATTCATG-----GCTCCTCCTCCT-----	573

Db 144 -----PheMetAsnGlyGlyLeuGlyAlaThrSerProLeuG1 156
QY 574 ----CATCTCTCCACTTCATCTTGGATCAITTAAGACACTATGATGTTCTCAAA 629
Db 156 yValHisProSerAlaGlnSerProMetGlnHisLeuGly
QY 630 CAACATGTGGGGTTTGAAGCAAAATAGTGAGTTTCAGGCAATTTTCAGGTGTAGTTGGTCC 689
Db 170 -----ValGlyMe 172
QY 690 AAGTGAACCATGATGCTACATTCGGTGAAGAAGATTTCGGTTCTTAATTTTGAATAA 749
Db 172 tGluAlaProLeuLeuGly-----PheProThrMetAsnSerAsnLe 186
QY 750 AAGAAACAATGAGCTTTCATGCTTTCAGTCTTCATGATGTTTCTGATGATGCTCGGAGAT 809
Db 186 userGluValGlnLysValLeuGlnIleValAspAsnThr----- 199
QY 810 AAGTCTTTGTGAGCTACAGATTAGCTCTAGAGCAAGCTTCTTCAGACACAAGACAT 869
Db 200 -----ValSerArgGlnLysMetAspCysLysThrGluAspI1 212
QY 870 TTCT-----AATAAGCTTGTACTCAAGTTTCTCTCACTAT 908
Db 212 eserLysLeuLysGlyTyrHisMetLysAspProCysSerGlnProGluGlnGlyVa 232
QY 909 ATTTGGCTCAAATACCTTCACTCTGTTCAAGAAATATCTATCTATTTTCGCGCATACTC 968
Db 232 lThrSerProAsnIleProProValGlyLeuProValSerHis----- 247
QY 969 GCTCATTTATTCATCTCAGGACCGAGTCTGAGTCTGATGCTTCTTACTTCTACG 1028
Db 248 -----AsnGlyAlaThrLysSerIleLeuAspTyrTh 258
QY 1029 TTTTGAGATATAACTGAG-----TTTCTGATGCTGATTTCTTAATAA 1070
Db 258 rleuGluLysValAsnGluAlaLysAlaCysLeuGlnSerLeuThrThrAspSer----- 276
QY 1071 CTCGAGGCGGGTTTCGATCTACATTTCAAGGAGAGCATTAGAAGCAAGAAACCA 1130
Db 277 -----ArgArgGlnIleSerAsnIleLysLysGly 287
QY 1131 TCTCTTGGATCTTCTCAAAATGGTG---GATGATCGATATAGTCTTCGTTAGATGAGAT 1187
Db 287 sleuArgThrLeuIleAspLeuValThrAspAspLys-----MetIleGluAs 303
QY 1188 TCATACGGTTATATCAGCGTTCCATGCT-----GCAACCGAGTTAGATCCA----- 1233
Db 303 nHisSerIleSerThrProPheSerCysGlnPheCysLysGluSerPheGlyProI1 323
QY 1234 -----CAGTTACACACCGGTTT-----GCCCTCCAAACCGTTTC 1268
Db 323 eProLeuHisGlnHisGluArgTyrLeuCysLysMetAsnGluGluIleLysAlaVal 343
QY 1269 CTTCTTATACAAGAACCTGAGAGAGAGA-----ATCTGCAAGAAGATAAT 1313
Db 343 uGlnProHisGluAsnIleValProAsnLysAlaGlyValPheValAspAsnLysAlaLe 363
QY 1314 CTCTATGGGATCTGATTGGAGAGGCAAGACAGACTCAAGAAACCTCTATGTTCCA 1373
Db 363 uLeuLeuSerSerValLeuSerGlu---LysGlyLeuThrSerProIleAsnProTyrLy 382
QY 1374 CCAGCATTCCTTCTCAGCAGCTGAAACGAAAGAACCATCAGATTTCGAGACCTCAACG 1433
Db 382 aspHis----- 384
QY 1434 AGGTTTGCCTGAGAAATCTGTTTCGGTTCTACGGAATTGGATTTCCAAAACCTTCCTTCA 1493
Db 385 -----MetSerValLeuLysAlaTyrTyrAlaMetAsnMet----- 396
QY 1494 CCTTACCCGAAAGATTCGGAACAACATCTTCTAGCTATACAGTCTGCTTGCAGAGAG 1553

Db 397 ---GluProAsnSerAspGluLeuLeuLysIleSerIleAlaValGlyLeuProGlnG1 415
QY 1554 TCAGGTATCAAACTGGTTTATAAATCGCGGGTTAGGCTATGGAAGCCGATGATAGAAGA 1613
Db 415 uPheValLysGluTrpPheGluGlnArgLysVal----- 426
QY 1614 GATGTATGCGAAATCAACAAGAGG-----AAGCTCAATPAACAGTCACATTC 1661
Db 427 ---TyrGlnTyrSerAsnSerArgSerProSerLeuGluArgThrSerLysProLeuAl 445
QY 1662 ACCAAC---GGACCAACTCTTCGA-----ATGCCAAATCT 1695
Db 445 aProAsnSerAsnProThrThrLysAspSerLeuLeuProArgSer 460
RESULT 9
US-09-457-040B-27
; Sequence 27, Application US/09457040B
; Patent No. 6387641
; GENERAL INFORMATION:
; APPLICANT: Vertex Pharmaceuticals Incorporated
; APPLICANT: Bellon, Steve
; TITLE OF INVENTION: Crystallized P38 Complexes
; FILE REFERENCE: VPI/98-14
; CURRENT APPLICATION NUMBER: US/09/457,040B
; CURRENT FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
; LENGTH: 1584
; TYPE: PRT
; ORGANISM: DICI - Dictyostelium Discoideum
US-09-457-040B-27
Alignment Scores:
Pred. No.: 0.0128 Length: 1584
Score: 109.50 Matches: 105
Percent Similarity: 29.98% Conservative: 68
Best Local Similarity: 18.20% Mismatches: 187
Query Match: 3.33% Indels: 217
DB: 4 Gaps: 27
US-09-423-575-1 (1-1886) x US-09-457-040B-27 (1-1584)
QY 292 CATAGAAACCAATGGACAAACAACAACAACACTTTTGTCTCTGGATAATGTC 351
Db 494 AsnAsnAsnAsnSerAsnSerAsnAsnAsnAsnAsnAsnAsnAsnSerAsnSer 513
QY 352 ATGATAACCAAAATCCTCTCTCATGATTTTATACCTCAAGAGAGATTCACTTCA 411
Db 514 AsnSerAsnAsnAsnIleAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnIle 533
QY 412 TTCTCAACATGCTTCCA-----TGGATACCATCAGATCCTCTACAATGGGT 465
Db 534 TyrLeuThrLysProSerIleGlySerThrAspGluSerThrGlySerLeuGly 553
QY 466 GGTCTTGATATTTTCAATCTCTATGCTGCTAACAATACTATCATCTCTCCACGGTCT 525
Db 554 Gly-----AsnAsnSerSerGlyAsnAsnSerSerSerGly----- 566
QY 526 ATCGATGTTCAAGATACCCCAATGTTGAGTTATGCTTCATCGCTCCTCTCTCAT----- 576
Db 567 ---SerIleGlyAsnAsnSerIleIleLysGlnArgSerProHisSerIleAsn 585
QY 577 ---CCTCCTCCATCTTCATCCTTTGGATCATTTAAGACACTATGATGATTCCTCAACAC 633
Db 586 GlyProLeuMetLeuProPro-----SerSerThrAsnAsn 597
QY 634 ATGTGGGGTTTGAAGCAAAATAGTGTTCAGGCATTT---TCAGGTGTAGTTGTGTC 690
Db 598 -----AsnAsnAsnIleTyrSerSerTyrAsnSerThrThrAlaGlySer 612
QY 691 AGTGAACCAATGATGCTCTACATTCGGTGAAGAGATTTCCTGTTTCTTAATTCGAATAA 750

Db 613 SerThrThrLeuProThrLeuAsnHisProIlePheGlyAsnThrThrSerAsn--- 631
 QY 751 AGAAACAATGAGCTTTCATAGTCTTCGATCAGAT-----GTTTCTGATGATGC 801
 Db 632 AsnAsnSerSerThrLeuSerValGlyGlyAsnAsnLeuLeuGlyArgHisCys 651
 QY 802 TCGGAGATAAGTCTTTGTCAGCTACAAGA----- 831
 Db 652 GlnSerLeuProIleThrAlaSerThrAsnHisThrLeuSerSerSerLeuGlyValSer 671
 QY 832 -----TTAGCTCTCAGAGCAAGCTTCTTCGACGACCAAGACATTTCTAAATAAGCTGTT 885
 Db 672 PheSerSerProSerSerSerProLysThrSerProArgLysIleValAsnSerSerGlu 691
 QY 886 ACTCAAGTTTC----- 897
 Db 692 AspLeuGlyPheValGlnThrPheGlnAspGlnAspGlyGlnProProSerAlaTrpArg 711
 QY 897 ----- 897
 Db 712 ArgCysGlyLysSerIleLysThrLysAspAspIleThrLeuThrIleLysLysLys 731
 QY 898 -----TCTCAACTTATATT 912
 Db 732 ThrSerValAlaMetAlaAspArgProPheSerSerAsnSerSerThrIleCysTyr 751
 QY 913 GGCTCAAAATACCTT-----CAC 930
 Db 752 PheGluValThrLeuGluGlyHisAspLysLysGlySerIleThrValGlyLeuSerHis 771
 QY 931 TCTGTTCAAGAAATATCTATCTCAATTCGCC-----GCATACTCGCTCGATTTATCTATCT 984
 Db 772 SerThrTyProPheIleLysHisIleGlyArgGluProLysSerTyArgPheSerSer 791
 QY 985 CGAGAACCGAGTCAGGAGCTGCTAGTTCAGCCCTTACTTCAGCTTTTGAATATACT 1044
 Db 792 GluGlyGluLysTyArgLysSer-----GluIleGly 802
 QY 1045 GAG-----TTTCTGTGTTGATTTCTAATAACTCGAGCGGGTTTC 1086
 Db 803 GluProTyGlyProPhePhePheAspGlyAspSerIleAlaSerSerCysValle 822
 QY 1087 GGA-----TCTACATTTCAAGAGGAGACATTAGAACAAAGAAACCCATCTCTTG 1137
 Db 823 GlyCysGlyIleAsnThrSerThrArgAspIlePhePheThrLysAsnGlyHisTyLeu 842
 QY 1138 GATCTTCTCAATGGTGGATGATGATATAGTATTGCGTAGATGAGATTCATACGGTT 1197
 Db 843 GlyVal----- 844
 QY 1198 ATATCAGCGTTTCCATGCTGCAACCGAGTTAGATCCACAGTTTACACCCCGTTTGCCTC 1257
 Db 845 -----AlaPheSerArgValThrSer---AspPro----- 853
 QY 1258 CAAACCGTTTCTCTTTATACAGAACCTGAGAGAGA-----ATCTGC 1302
 Db 854 -----LeuTyProSerIleSerPheArgGlyValValGlyGlyLeuCys 868
 QY 1303 AAGAGATAATCTCTATCGGATCTGTATTGGAGAGAGCAAGACAAGACTCAAGAAACC 1362
 Db 869 ValAlaThrPheProGlyGly----- 875
 QY 1363 TCTATGTTCCACGAGCTGCTTCTTTCAGAGCTGAAACAAAGAACCATCAGATTGG 1422
 Db 876 -----HisPheArgPheAsnIleGluAspLeuProGlyIleSerProSerValTrp 892
 QY 1423 AGACCTCAACGAGGTTTCCCTTACAGAAATCTGTTTCGGTTCTACCGAATTTGG----- 1473
 Db 893 ThrGluAlaLeuGly---ProAspArgGlnGlySerGlyPheLysAsnTrpAlaProAsn 911
 QY 1474 -----ATGTTCCAAACTTCTCTTCCAC---CCTTACCCGCAAGAT----- 1509
 Db 912 AspValAlaIleTrpLeuGluSerPheAsnTyArgGlnTyArgLysAsnPheArgAsp 931

QY 1510 -----TCGAGAAACATCTTTAGCTATACGAGTGGTTCACAAAGATCAGGTA 1560
 Db 932 AsnAsnIleSerGlyArgHisLeu-----GluGlyIleThrHisAlaMetLeu 947
 QY 1561 TCAAACTGGTTT-----ATAAAT 1578
 Db 948 LysAsnAspLeuGlyIleGluProTyGlyHisArgGluAspIleLeuAsnArgLeuAsn 967
 QY 1579 GCGCGGTTAGGCTATGAGGAGCGGATGATAGAGAGATCTATCGGAAATG 1629
 Db 968 ArgMetIleGlnIleTrpAsnAspLysSerProAspSerTyProLysIle 984
 RESULT 10
 US-09-107-532A-3678
 ; Sequence 3678, Application US/09107532A
 ; Patent No. 6583275
 ; GENERAL INFORMATION:
 ; APPLICANT: LYNN A Doucette-Stamm and David Bush
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 ; NUMBER OF SEQUENCES: 7310
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 ; STREET: 100 Beaver Street
 ; CITY: Waltham
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02354
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD/ROM ISO9660
 ; COMPUTER: PC
 ; OPERATING SYSTEM: <Unknown>
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/107,532A
 ; FILING DATE: 30-Jun-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/085,598
 ; FILING DATE: 14 May 1998
 ; APPLICATION NUMBER: 60/051571
 ; FILING DATE: July 2, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ariniello, Pamela Deneka
 ; REGISTRATION NUMBER: 40,489
 ; REFERENCE/DOCKET NUMBER: GTC-012
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (781)893-5007
 ; TELEFAX: (781)893-8277
 ; INFORMATION FOR SEQ ID NO: 3678:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 522 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Enterococcus faecium
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (B) LOCATION 1...522
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 3678:
 US-09-107-532A-3678
 Alignment Scores:
 Pred. No.: 0.00993 Length: 522
 Score: 108.00 Matches: 86
 Percent Similarity: 38.84% Conservative: 81
 Best Local Similarity: 20.00% Mismatches: 149
 Query Match: 3.29% Indels: 114
 DB: 4 Gaps: 22
 US-09-423-575-1 (1-1886) x US-09-107-532A-3678 (1-522)

427 CCATGAATACCATCAGATCAGATCCTCTCAAAATGGGTGGTCTTGATATTTCATTTCT 486
Db |||||
43 ProTtp-----GlyTtpProLeuIleGlyHis 51
QY |||||
487 ATGCTGACT-----AACAAATACTATCATCTTCTCCACGGTCTCATGATGTTCAA 537
Db |||||
52 MetLeuThrLeuGlyLeuAsnProHisLeuAlaLeuSerArgMetSerGlnGlnTyrGly 71
QY |||||
538 GATAACCGCAATGTTGAGTTCATGCTCTCTCTCTCATCTCTCCACCTTCATCTTTG 597
Db |||||
72 AspValLeuGlnIleArgIleGlyThrPro-----ValValValLeuSerGlyLeu 88
QY |||||
598 GATCATTTAAGACACTAT-----GATGATTCTCTCAAAACAATGTGGGT 642
Db |||||
89 AspThrIleArgGlnAlaLeuValArgGlnGlyAspAsp-----SerAsnGlyGln 101
QY |||||
643 TTTGAAGCAATAGTGGTTCAGGCAATTTTCAGGTGAGTGTGCTCCAGGTGAACCAATG 702
Db |||||
102 PheLysGlyArgProAspLeuTyrThrPheThrLeuIle-----SerAsnGlyGln 118
QY |||||
703 ATGTCATCATGCTGGAAGAAGATTTCCTGTTCTTAATTTCCGAATAAAGA----- 753
Db |||||
119 SerMetSerPheSerProAspSerGlyProValTrpAlaAlaArgArgLeuAlaGln 138
QY |||||
754 RACATGAGCTTTCATTCAGTCTTCATCAGATGTTCTGATGATGCTCGAGATAGT 813
Db |||||
139 AsnGlyLeuLysSerPheSerIleAlaSerProAlaSerThrSer----- 155
QY |||||
814 CTTTGTGAGCTTACAGATTAGCTTCAGAGCAAGCTTCTTCAGCAGCAAAAGACATTCT 873
Db |||||
156 ---CysTyrLeuGluGluHisValSerLysGluAlaGluValLeuIleSerThrLeuGln 174
QY |||||
874 AATAAGCTT-----GTTACT 889
Db |||||
175 GluLeuMetAlaGlyProGlyHisPheAsnProTyrArgTyrValValValSerValThr 194
QY |||||
889 CAAGGTTCTCTCAACTATATTGGCTCAAAATACCTTCACTCTGTTCAGAAATACTA 948
Db |||||
195 AsnValIleCysAlaIleCysPheGlyArgTyrAspHisAsnHisGlnGluLeuLeu 214
QY |||||
949 TCTATTTCCGCGATATCTGCTGATTTATCTCTGAGGAACCGAGTCAGAGTGCT 1008
Db |||||
215 Ser-----LeuValAsnLeuAsn---AsnAsnPheGlyGluValValGlySerGly 230
QY |||||
1009 AGTTCCAGCTTTTACTTCCAGCTTTGAGATATACTGATGTTCTTGTGTTGATTCTAAT 1068
Db |||||
231 AsnProAla-----GluPheIleProIleLeuArgTyrLeuProAsnProSerLeu 247
QY |||||
1069 AACTCG-----GAGCGCGGTTTCGATCTACATTTCAGAGGAGAGCATTAGAA 1116
Db |||||
248 AsnAlaPheLysAspLeuAsnGlnLysPheTyrSerPheMetGlnLysMetValLysGlu 267
QY |||||
1117 GCAAGRAAAC-----CATCTCTGATCTTCTTCAATGGTGGATGATCGA 1164
Db |||||
268 HisTyrLysThrPheGluLysGlyHisIleArgAspIle-----ThrAspSerLeu 284
QY |||||
1165 TATAGTCATTGC-----GTAGATGAG----- 1185
Db |||||
285 IleGluHisCysGlnGlnLysGlnLeuAspGluAsnAlaAsnValGlnLeuSerAspGlu 304
QY |||||
1186 -----ATTATACGTTATATACGCTTCATCTGCAACCGAGTAGATCCACAGTTA 1239
Db |||||
305 LysIleIleAsnIleValLeuAspLeuPheGlyAlaGlyPheAsp----- 319
QY |||||
1240 CACACCGGTTGCGCTCCAAACGTTCTTCTTATACAGAAGACCTGAGAGAGAGATC 1299
Db |||||
320 ---ThrValThrAlaIleSerTrpSerLeuMetTyrLeuValMetAsnProArgVal 338
QY |||||
1300 TGCAGAGAGATA---ATCTTATGGATCTCTATTGGAGAGAGCAAGACAGACTCAA 1356
Db |||||
339 GlnArgLysIleGlnGluLeuAspThrValIleGlyArgSerArg----- 354
QY |||||
1357 GAAACCTCTATGTTCCACCAAGATTGCTCTTCTTCAGCAGCTGAAACGAAAGAACCATCAG 1416

355 -----ArgProArgLeuSerAspArgSerHis--- 363
Db |||||
1417 ATTTGGAGACCTCAACGAGGTTTGCTGAGAAATCTGTTTCGTTCTACGGAATTGGATG 1476
QY |||||
364 -----LeuProTyrMetGluAlaPheIleLeuGluThrPheArg 376
Db |||||
1477 TTCCAAATCTTCCT-----CACCCCTTACCCGAAAGATTTCGAGAAACAT 1521
QY |||||
377 HisSerSerPheValProPheThrIleProHisSerThrThrArgAspThrSerLeuLys 396
Db |||||
1522 CTTCTAGCTATACGAAGTGGCTTGACAGAAGTCAGGTATCAAACTGGTTTATAATCGG 1581
QY |||||
397 GlyPheTyrIleProLysGlyArgCys---ValPheValAsnGlnTrpGlnIleAsnHis 415
Db |||||
1582 CGGGTTAGGCTATGG 1596
QY |||||
416 AspGlnLysLeuTrp 420
Db |||||
RESULT 12
US-08-480-662-2
; Sequence 2, Application US/08480662
; Patent No. 5759782
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira
; APPLICANT: Brinkmann, Ulrich
; TITLE OF INVENTION: CELLULAR APOPTOSIS SUSCEPTIBILITY PROTEIN (CSP) AND AN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,662
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NH112.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 971 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; US-08-480-662-2
Alignment Scores:
Pred. No.: 0.0207 Length: 971
Score: 106.50 Matches: 117
Percent Similarity: 34.05% Conservative: 74
Best Local Similarity: 20.86% Mismatches: 193

```

Query Match: 3.24% Indels: 177
DB: Gaps: 26
US-09-423-575-1 (1-1886) x US-08-480-662-2 (1-971)
QY 181 CTTGATGTTTAAACACATTTTGGAAATTTGATGTAATAATTTCTTTGGAAAGTTGT 240
D 8 LeuGlnThrLeuThrGluThrLeu-----LysLysThrLeuAsp----- 20
QY 241 GTTGCTGAATCTCCCAAGGTTCTATCAGAAAGAGAGTAAGATTTTCATAGAAAC 300
D 21 -----ProAspProAlaIleArg-gProAlaGluLysPhe----- 32
QY 301 CCAATGGACACAAACAAACAAACACTTTTAGTTCTCTGCAATAATGTCTAGCTAAAC 360
D 33 -----LeuGluSerValGluGlyAsn 39
QY 361 CAAAT-----CCTCTTCTCATGATTTTATACCTTCAAGAGAAGAT----- 402
D 40 GlnAsnTyProLeuLeuLeuLeuThrLeuLeuGluLysSerGlnAspAsnValIleLys 59
QY 403 -----TCAACTTCTCAATCTCAACATGCTTCCA-----TGAATACCATCAGATCA 447
D 60 ValCysAlaSerValThrPheLysAsnTyIleLysArgAsnTrpArgIleValGluAsp 79
QY 448 GATCCTCTCAAAATGGGTGGCTTTGAT-----ATTTCAATTTCT 486
D 80 GluProAsnLysIleCysGluAlaAspArgValAlaIleLysAlaAsnIleValHisLeu 99
QY 487 ATGCTGACTAAACATATTATCATTCTCTCCACGGTCTATCGATGTTCAAGATAACCGC 546
D 100 MetLeu-----SerSerProGluGlnIleGlnLysGlnLeuSerAsp 113
QY 547 AATGTTGAGTTGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 606
D 114 AlaIleSerIleIleGlyArgGluAspPhePro----- 124
QY 607 AGACTATGATGATCTCTCAACAAACATGCTGGGTTTGAAGCAAAATAGTAGTTTCAG 666
D 125 GlnLysTrpProAspLeuLeuThrGluMetValAsnA-gPheGlnSerGlyAspPheHis 144
QY 667 GCATTTTTCAGGTGATGTTGGTCCAGTGAACCAATGATGTCACATTTGGTGAAGAAGAT 726
D 145 ValIleAsnGlyValLeuArgThrAlaHisSerLeuPheLysArgTyArgHisGlu--- 163
QY 727 TTCCTGTTCTAATTTGGAATAAAAGAAACATGAGTTTCAATGATGTTTCATCAGAT 786
D 164 -----PheLysSerAsnGluLeuTrpThrGluIleLysLeuValLeu----- 177
QY 787 GTTCTGTGATGATGCTCGAGATAAGTCTTTGTCAGCTACAGATTAGCTCAGACAA 846
D 178 -----AspAlaPheAlaLeuProLeuThrAsnLeuPheLysAlaThrIleGluLeu 194
QY 847 GCTTCTTGACGACGACAAAGACATTTCTAATAACGTTGTACTCAAGGTTTCTCTCAACT 906
D 195 CysSerThrHisAlaAsnAspAlaSerAlaLeuArgIleLeu-----PheSerSerLeu 212
QY 907 ATATTTGGCTCAAAATACCTTCACTCTGTTCAAGAAATACTATCTCAATTCGCCCATAC 966
D 213 IleLeuIleSerLysLeuPheTySerLeu----- 222
QY 967 TCGCTCGATTATCACTCGAGGAACCGAGTCAGGAGCTGCTAGTTTCAGCCCTTACTTCA 1026
D 222 ----- 222
QY 1027 CGTTTGTGAATATACTAGTTTCTTGATGTTGATCTTAATACTCGGAGCGGGTTTC 1086
D 223 AsnPheGlnAspLeuProGluPheTrpGluGlyAsnMetGluThrTrpMetAsnAsnPhe 242
QY 1087 GGATCT-----ACATTTCAAAGGACGATTAGAGCA-----AAGAAACCCAT 1131
D 243 HisThrLeuLeuThrLeuAspAsnLysLeuLeuGlnThrAspAspGluGluAlaGly 262

```

```

RESULT 13
US-08-918-190-2
; Sequence 2, Application US/08918190
; Patent No. 6072031
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ita
; APPLICANT: Brinkmann, Ulrich
; TITLE OF INVENTION: CELLULAR APOPTOSIS SUSCEPTIBILITY
; TITLE OF INVENTION: PROTEIN (CSP) AND ANTISENSE CSP
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson and Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/918,190
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/480,662
; FILING DATE: 07-JUN-1995

```

```

QY 1132 CTCTTGATCTTCTT-----CAAATGGTG 1155
D 263 LeuLeuGluLeuLeuLysSerGlnIleCysAspAsnAlaLeuTyAlaGlnLysTy 282
QY 1156 GATGATCATATATGATCGTAGATGATGATGATGATGATGATGATGATGATGATGATG 1215
D 283 AspGluLeuPheGlnArgTyLeuProArgPheValThrAlaIleTrpAsnLeuVal 302
QY 1216 GCACACC-----GAGTTAGATCCACAGTTACACACCGGTTTCCCTCCAAACCGTTCC 1269
D 303 ThrThrGlnGluValLysTyAspLeuLeuValSerAsnAlaIleGlnPheLeuAla 322
QY 1270 TTCTTA-----TACAAGAACCTCGAGAG-----AGA 1296
D 323 SerValCysGluArgProHisTyLysAsnLeuPheGluAspGlnAsnThrLeuThrSer 342
QY 1297 ATCTGCAAGAGATATCTCTATGGGATCTGTATTGGAG-----GAGCAAA 1344
D 343 IleCysGlnLysValIle-ValProAsnMetGluPheArgAlaAlaAspGluAla-p 362
QY 1345 GACAAGACTCAAGAAACCTCTATGTTCCACAGCATTTGCCTTCTTCCAGCAGCTGAAACGA 1404
D 362 heGluAspAsnSerGluGluTyIleArgArgAspLeu----- 374
QY 1405 AAGAACCATCAGATTTGGAGACCTCAACGAGGTTTGCCTGAGAAATCTGTTTCGTTCTA 1464
D 375 --GluGlySerAspIleAspThrArgArgArgAlaAlaCysAspLeuValArgGlyLeu 394
QY 1465 CGGAATTTGGATGTTCCAAAACCTCTTCCCTTACCCGAAAGATTC---GGAGAAACAT 1521
D 394 yLysPhe-----PheGluGlyProValThrGlyIlePheSerGlyTyVala 410
QY 1522 CTCTCT-----AGCTATACGAAG-----TGCTTGACCAAGAGTCAG 1557
D 410 snSerMetLeuGlnGluTyAlaLysAsnProSerValAsnTrpLysHisLysAspAla 430
QY 1558 GTATCAACTGGTTTATAAATGCGCGGTTAGGCTATGGAAGCCGATGATAGAGATG 1617
D 430 laIleTyLeuValThrSerLeuAla----- 438
QY 1618 TATCGGAAATGAACAAGAGAGCTCAATAACAGTCACATTCACCCCAACCGACCAAC 1676
D 439 -----SerLysAlaGlnThrGlnLysHisGlyIleThrGlnAlaAsn 452

```

ATTORNEY/AGENT INFORMATION:

NAME: Israel, Ned A
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NH112.001A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-235-8550
 TELEFAX: 619-235-0176
 TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
 LENGTH: 971 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 ORIGINAL SOURCE:

US-08-918-190-2

Alignment Scores:

Pred. No.: 0.0207 Length: 971
 Score: 106.50 Matches: 117
 Percent Similarity: 34.05% Conservative: 74
 Best Local Similarity: 20.86% Mismatches: 193
 Query Match: 3.24% Indels: 177
 DB: 3 Gaps: 26

US-09-423-575-1 (1-1886) x US-08-918-190-2 (1-971)

```

181 CTTCACTTTTAAACACACATTTTGGAAATTTTGATGTAAATAATCTCTTTGGACGTTGT 240
182 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
183 8 LeuGlnThrLeuThrGluThrLeu-----LysLysThrLeuAsp----- 20
184
241 GTTGCTGAAATCTCCCAAGGTTCTATCAGAGAAGAGGATAAATTTCATAGAAC 300
242 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
243 21 -----ProAspProAlaIleArgProAlaGluLysPhe----- 32
244
301 CCAATGGACACAAACACACACACTTTTGTCTCTGGATAATGCTCATGACTAAC 360
302 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
303 33 -----LeuGluSerValGluLysAsn 39
304
361 CAAAT-----CCTCTTCTCATGGATTATATACCTTCAAGAGAGAT----- 402
362 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
363 40 GlnAsnTyrProLeuLeuLeuThrLeuLeuGluLysSerGlnAsnValIleLys 59
364
403 -----TCAACTTCATCTCAACATGCTTCCA-----TGAATACCATCATCA 447
404 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
405 60 ValCysAlaSerValThrPheLysAsnTyrIleLysArgAsnTrpArgIleValGluAsp 79
406
448 GATCCTCTCAAAATGGGTGGCTTGTAT-----ATTTTCAATTCT 486
449 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
450 80 GluProAsnLysIleCysGluAlaAspArgValAlaIleLysAlaAsnIleValHisLeu 99
451
487 ATGCTGCACTAAATAATCTATCTCTCTCCACGGTCTATCGATGTTCAAGATAACGC 546
488 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
489 100 MetLeu-----SerSerProGluGlnIleGlnLysGlnLeuSerAsp 113
490
547 AATGTTGAGTTCATGGCTCCTCCTCTCATCTCCTCCATCTCATCTTTGGATCAATTA 606
548 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
549 114 AlaIleSerIleIleGlyArgGluAspPhePro----- 124
550
607 AGACACTATGATGATCTCTCAACACATGTGGGGTTTTGAAGCAATAGTAGTTTCAG 666
608 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
609 125 GlnLysTyrProAspLeuLeuThrGluMetValAsnArgPheGlnSerGlyAspPheHis 144
610
667 GCATTTTCAAGGTAGTTGGTCCCAAGTGAACCAATGATGTCTACATCTCGGTGAAGAAGAT 726
668 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
669 145 ValIleAsnGlyValIleuArgThrAlaHisSerLeuPheLysArgTyrArgHisGlu 163
670
727 TTCCCGTTTCTTAATTCGAATAAAGAACAAATGAGCTTTTCATTGAGCTTTGCAATCAGAT 786
728 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 14

US-09-234-232-2

; Sequence 2, Application US/09234232

Qy	1087	GGATCT	-----ACATTTC	AAAGAGACG	CATTAGA	CAACCA	-----	AGAAACCCAT	1131
		:::	:::	:::	:::	:::		:::	
Db	243	HisThrLeu	ThrLeu	AspAsn	LysLeu	GlnThr	AspGlu	GluGlu	AlaGly 262
Qy	1132	CTCTTGAT	CTCTT	-----	-----	-----	-----	CAATGGTG	1155
Db	263	LeuLeu	GlnLeu	LeuLys	SerGln	IleCys	AspAsn	AlaLeu	TyrAlaGlnLysTyr 282
Qy	1156	GATCAT	CGATAT	GATCTCAT	TGCGTAGAT	GAGATT	CATACGG	TTATATCAG	CGGTTCCTCATGCT 1215
				:::	:::				
Db	283	AspGlu	GluPhe	GlnArg	TyLeu	ProArg	PheVal	ThrAla	IleTrpAsnLeuVal 302
Qy	1216	GCAACC	-----	-GAGTTAGAT	CCACAGCT	TACACAC	CCGGTTG	TGCCCTCC	CAAAACCGTTTCC 1269
Db	303	ThrThr	GlyGln	GluVal	LysTyr	AspLeu	LeuVal	SerAsn	AlaIleGlnPheLeuAla 322
Qy	1270	TTCTTA	-----	-----	-----	-----	-----	-----	AGA 1286
		:::	:::	-----	TACAGAA	CACTGAG	AGAG	-----	AGA 1286
Db	323	SerVal	CysGlu	ArgPro	HisTyr	LysAsn	LeuPhe	GluAsp	GlnAsnThrLeuThrSer 342
Qy	1297	ATCTG	CAAGAG	ATAAT	CTCTAT	CGGAT	CTCTAT	TGGAGA	-----GAGGCAAA 1344
Db	343	IleCys	GluLys	ValIle	-ValPro	AsnMet	GluPhe	ArgAla	AlaAspGluGluAla-P 362

223	AsnPheGlnAspLeuProGluPheTrpGluGlyAsnMetGluThrTyrPheMetAsnAsnPhe	242
1087	GGATCT-----ACATTTCAAAGGAGAGCATTAGAAGCA-----AAGAAAACCCAT	1131
	::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::	
243	HlsThrLeuLeuThrLeuAspAsnLysLeuLeuGlnThrAspGluGluGluAlaGly	262
1132	CCTCTTGGAATCTCTT-----CAAATGGTG	1155
	:::: :::: :::: :::: :::: ::::	
263	LeuLeuGluLeuLeuLysSerGlnIleCysAspAsnAlaAlaLeuTyrAlaGlnLysTyr	282
1156	GATGATCGATATAGTCATTGCGTGATGATGAGATTCATCGGTTATATCATAGCGTTCATGCT	1215
	:::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::	
283	AspGluGluPheGlnArgTyrLeuProArgPheValThrAlaIleTrpAsnLeuLeuVal	302
1216	GC AAC-----GAGTTAGATCCACAGTTTACACCCGGTTGCCCTCCAAACCGTTCC	1269
	:::: :::: :::: :::: :::: ::::	
303	ThrThrGlyGlnGluValLysTyrAspLeuLeuValSerAsnAlaIleGlnPheLeuAla	322
1270	TTCCTTA-----TACAAGAACCTCGAGAG-----AGA	1296
	::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::	
323	SerValCysGluArgProHisTyrLysAsnLeuPheGluAspGlnAsnThrLeuThrSer	342
1297	AICTGCAAGAAGATAATCTCTATGGGATCTGTATTGGAGA-----GAGGCAAA	1344
	:::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::	
343	IleCysGluLysValIle-ValProAsnMetGluPheArgAlaAlaAspGluGluAla-P	362
1345	GACAAAGACTCAAGAAAACCTCTATGTTCCACCAGCATTTGCCCTTCTTCCAGCAGCTGAACGA	1404
	:::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::	
362	heLuAspAsnSerGluGluTyrlleArgArgAspLeu-----	374
1405	AAGAACATCAGATTGGAGACCTCAACGAGTTTCCTCAGAAATCTGTTTCGGTTCTA	1464
	:::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::	
375	--GluGlySerAspIleAspThrArgArgargalalaCysAspLeuValArgGlyLeuC	394
1465	CGGAATTGGATGTTCCA AAACTCCTTCACCTTACCCGAAAGATTCT---GGAGAAACAT	1521
	:::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::	
394	ysylsPhe-----PheGluGlyProValThrGlyllePheSerGlyTyrVaia	410
1522	CTTCT-----AGCTATACGAG-----TCGCTTGCAAGAAGTCAG	1557
	:::: :::: :::: :::: :::: ::::	
410	snSerMetLeuGlnGluTyrAlaLysAsnProSerValAsnTrpLysHlsLysAspAla	430
1558	GTATCAAACTGGTTTTAAAAATGCGCGGTTTAGGCTATGGAAGCCGATGATAGAGAGATG	1517
	:::: :::: :::: :::: :::: ::::	
430	laileTyrLeuValThrSerLeuAla-----	438
1618	TATCGGAAGATGAACAAGAGAGAGCTCAATAACAGTCACTCAATCAACCCACGACCAAC	1676
	:::: :::: :::: :::: :::: ::::	
439	-----SerlysAlaGlnThrGlnLysHlsGlylleThrGhAlaAsn	452

Search completed: September 2, 2004, 12:14:17
Job time : 64.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: September 2, 2004, 12:11:49 ; Search time 332.5 Seconds
(without alignments)
3574.213 Million cell updates/sec

Title: US-09-423-575-1
Perfect score: 3284
Sequence: 1 attatgtataaaattgtgc.....ttaaaaaaaaaaaaaaaaaa 1886

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1298764 seqs, 315065143 residues

Total number of hits satisfying chosen parameters: 2597529

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09423575/runat_01092004.161109.18704/app.query.fasta.1.2055
-DB=PublishedApplications_AA -QWMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09423575@cgn.1.161@runat_01092004.161109.18704
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1
US-10-424-599-273470
; Sequence 273470, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424.599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 273470
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_88966C.1.pep
US-10-424-599-273470
Alignment Scores:

1	503	15.3	702	12	US-10-424-599-273470	Sequence 273470, A
2	473.5	14.4	550	12	US-10-425-114-37053	Sequence 37053, A
3	472	14.4	680	12	US-10-225-066A-336	Sequence 336, App
4	472	14.4	680	15	US-10-374-780A-2672	Sequence 2672, App
5	467.5	14.2	475	12	US-10-425-114-58821	Sequence 58821, A
6	467.5	14.2	499	12	US-10-425-114-68801	Sequence 68801, A
7	467.5	14.2	584	16	US-10-437-963-122156	Sequence 122156, A
8	465	14.2	618	16	US-10-437-963-110193	Sequence 110193, A
9	454	13.8	678	16	US-10-437-963-162637	Sequence 162637, A
10	443	13.5	538	12	US-10-424-599-180416	Sequence 180416, A
11	440.5	13.4	481	12	US-10-425-114-43080	Sequence 43080, A
12	439	13.4	642	16	US-10-437-963-134068	Sequence 134068, A
13	437	13.3	335	12	US-10-425-114-48619	Sequence 48619, A
14	431.5	13.1	642	16	US-10-437-963-143490	Sequence 143490, A
15	428.5	13.0	133	12	US-10-424-599-193114	Sequence 193114, A
16	428	13.0	447	12	US-10-425-114-41924	Sequence 41924, A
17	427	13.0	447	12	US-10-424-599-208750	Sequence 208750, A
18	426.5	13.0	884	16	US-10-437-963-170412	Sequence 170412, A
19	424	12.9	739	12	US-10-425-068-236	Sequence 236, App
20	424	12.9	739	15	US-10-374-780A-2330	Sequence 2330, App
21	424	12.9	739	15	US-10-425-114-42253	Sequence 42253, A
22	422	12.9	444	12	US-10-424-599-198122	Sequence 198122, A
23	422	12.9	463	12	US-10-425-114-40365	Sequence 40365, A
24	422	12.9	475	12	US-10-424-599-274000	Sequence 274000, A
25	421.5	12.8	461	12	US-10-437-963-117364	Sequence 117364, A
26	418	12.7	1123	16	US-10-425-114-39647	Sequence 39647, A
27	415.5	12.6	610	12	US-10-412-699B-270	Sequence 270, App
28	412.5	12.6	610	15	US-10-225-068-56	Sequence 56, Appl
29	412.5	12.6	610	15	US-10-374-780A-2332	Sequence 2332, App
30	412.5	12.6	610	15	US-10-424-599-206716	Sequence 206716, A
31	411.5	12.5	442	12	US-10-425-114-37177	Sequence 37177, A
32	410.5	12.5	442	12	US-10-412-699B-638	Sequence 638, App
33	408	12.4	627	12	US-10-425-066A-820	Sequence 820, App
34	408	12.4	627	12	US-10-286-264-18	Sequence 2482, App
35	408	12.4	627	15	US-10-374-780A-2482	Sequence 282020, A
36	407.5	12.4	654	12	US-10-424-599-282020	Sequence 108604, A
37	407.5	12.4	601	16	US-10-437-963-108604	Sequence 57007, A
38	386.5	11.8	377	12	US-10-425-114-57007	Sequence 137941, A
39	384	11.7	603	16	US-10-437-963-137941	Sequence 37932, A
40	383.5	11.7	224	12	US-10-425-114-37932	Sequence 284132, A
41	379	11.5	374	12	US-10-424-599-284132	Sequence 47120, A
42	378	11.5	374	12	US-10-425-114-47120	Sequence 284018, A
43	378	11.5	374	12	US-10-424-599-284018	Sequence 39635, A
44	370	11.3	479	12	US-10-425-114-39635	
45	359.5	10.9	385	12		

```

QY 1324 TCTGTATTGGAGAGAGGCAAAAGC-----AAGACTCAAGAAAACCTCTATG 1366
Db 339 LysThrLeu-----GlyGluAspAspCysLeuGlyValGlySerArgLeu 356
QY 1369 -----TTCACACAGCATTCCTTCAGCAGCTGAAACGAAAGAAC 1410
Db 357 ArgTyrValAspHisLeuArgGlnInArgAlaLeuGlnInLeuGlyMetIleGln 376
QY 1411 CATCAGATTGGAGACCTCAACAGGTTTCCTCGAGAAATCTGTTTCGGTTCCTACGGAAT 1470
Db 377 ProAsnAlaTrpArgProGlnArgGlyLeuProGluArgAlaValSerIleLeuArgAla 396
QY 1471 TGAATGTTCCAAAACCTTCCTTCCACCTTACCCCGAAAGATTCGGAGAAACATCTTCAGCT 1530
Db 397 TrpLeuPheGluHisPheLeuHisProTyrProLeuAspSerAspLeuValMetLeuAla 416
QY 1531 ATACGAAGTCGCTTGACAAAGATCAGGTATCAAACTGGTTTATAATGCGCGGGTTAGG 1590
Db 417 LysGlnThrGlyLeuAlaArgSerGlnValSerAsnTrpPheIleAsnAlaArgValArg 436
QY 1591 CTATGGAAGCCGATGATAGAGAGATGTATGCGGAAATGAACAAGAGGAGAGCTCAATAAC 1650
Db 437 LeuTrpIysProMetValGluGluMetTyrLeuGluGluIleLys----- 451
QY 1651 AGTCACATTCAACCCCAACGGA 1671
Db 452 GluHisGluGlnGlyAsnGly 458

RESULT 2
US-10-425-114-37053
; Sequence 37053, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 37053
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3060-003-B11_FLI pep
US-10-425-114-37053

Alignment Scores:
Pred. No.: 1,7e-37 Length: 550
Score: 473.50 Matches: 128
Percent Similarity: 47.50% Conservative: 62
Best Local Similarity: 32.00% Mismatches: 119
Query Match: 14.42% Indels: 91
DB: 12 Gaps: 13

US-09-423-575-1 (1-1886) x US-10-425-114-37053 (1-550)

QY 550 GTTAGTTCATGCTCTCTCTCTCATCTCTCCATCTCTCCATCTCTTCGATCATTTAAGA 609
Db 85 LeuGlnValGlnAlaProPro----- 92
QY 610 CACTATGATGATTCCTCAAAACAACATGTGG----- 639
Db 93 -----SerGlyTyrAsnLeuTrpThrProAlaMetAlaGlyValAlaAsp 108
QY 640 ---GGTTTGAAGAAAATGATGAGTTTCAGGCATTTTCAGGTGATGTGTCCTCAAGTGAA 696
Db 640 ---GGTTTGAAGAAAATGATGAGTTTCAGGCATTTTCAGGTGATGTGTCCTCAAGTGAA 696

```

```

Db 109 MetSerProGlnAlaGlnThrProGlyAlaAlaGlyAlaValSerAla 128
QY 697 CCAATGATGCTACATTCGGTGAAGAAGATTCCTCCGTTCTTAATTCGAATAAAGAAAC 756
Db 128 -----
QY 757 AATGAGCTTTTCATTCAGTCTTCATCAGATGTTCTGATGATGCTCGAGATAAGTCTT 816
Db 129 ---ValLeuSerLeuSer-----SerArgGluAlaProProValThrVal 142
QY 817 TGTGAGCTACAGATTAGCTCAGAGCAAGTCTTCGAGC-----AGCAAGACAT 870
Db 143 AlaAlaVal-----AlaAlaCysThrAspGluGlyLysTyrLeu 155
QY 871 TCTAATAAGCTTCTACTCAAGGTTTCTCTCAACTATATTTGGCTCAAAATACCTTCAC 930
Db 156 GlyValSerAlaThrProGlnGly-----GlnMetValMetSerSerLysTyrLeuLys 173
QY 931 TCTGTTCAAGAAATATCTATCTCATTCGCGGATCTCGCTCGATTTATTCATCTCGAGA 990
Db 174 AlaAlaGlnGluLeuLeuAspGluValValSerLysGlyValGluAspAlaLys 193
QY 991 ACGGAGTACGAGCTGCTAGTTCAGCCTTTACTTCA---CGTTTGAAGATAACTGAG 1047
Db 194 ThrAlaAlaAlaAlaThrLysSerLeuAlaAlaValLysLysGluAspSerGlu 213
QY 1048 TTTCTTGTAGTGT-----GATTTCTAATACTCGAGCGGTTTCGGA----- 1089
Db 214 GlyAlaSerGlyGlyGlyThrAspAspGlyAlaGlyAlaLysSerGlyGlyAlaAla 233
QY 1090 -----TCTACATTTCAAGGAGGATTTAGAGCAAGCAAGAAACCCATCTCTGGAT 1140
Db 234 ProGluMetSerThrAlaGluArgGlnMetGlnMetLysLysSerLysLeuLeuAsn 253
QY 1141 CTTCTTCAATGATGATGATATAGTATCTATTCGTTAGATGAGATTCATACGTTATA 1200
Db 254 MetLeuAspGluValGluGlnArgTyrArgGlnTyrHisGlyGlnMetGlnAlaValSer 273
QY 1201 TCAGCGTTCCATCTCTCAACCGAGTTA---GATCCACAGTTATACACCGGTTCCCTC 1257
Db 274 SerSerPheGluAlaAlaAlaGlyAlaGlySerAlaArgThrTyrThrAlaLeuAlaLeu 293
QY 1258 CAACCGTTTCTCTTATACAGACCTGAGAGAGATCTGCAAGAGATATCTCT 1317
Db 294 ArgThrIleSerArgGlnPheArgCysLeuArgAspAlaIleAlaSerGlnValArgAla 313
QY 1318 ATGGGATCTGTATTTGAGAGAGCAAGACAGACTCAAGAAACCTCT----- 1365
Db 314 AlaSerArgAlaLeu-----GlyGluAspAlaAspAlaAlaValAlaGlyArgThr 331
QY 1366 -----ATGTTCCACGAGATTCGCTTCTTCAGAG 1395
Db 332 ValGlySerArgLeuArgTyrIleAspHisGlnLeuArgGlnAlaLeuGlnGln 351
QY 1396 CTGAACGAAAGAACCATCAGATTTGGAGACCTCAACGAGGTTTGGCTGAGAAATCTGTT 1455
Db 352 LeuGlyMetMetGlnGlyAlaTyrArgProGlnArgGlyLeuProGluArgSerVal 371
QY 1456 TCGGTTCTACGAAATTTGATGTTCCAAATCTCTTCACCTTACCCGAAAGATTCGAG 1515
Db 372 SerIleLeuArgAlaTyrPheGluHisPheLeuHisProTyrProLysAspSerAsp 391
QY 1516 ARAATCTTCTAGCTATACGAGTGGCTTGACAAGAGTCAGGTATCAACTGGTTTATA 1575
Db 392 LysIleMetLeuAlaLysGlnThrGlyLeuThrArgSerGlnValSerAsnThrPheIle 411
QY 1576 AATGCGCGGTTTAGGCTTGAAGCCGATGATGAGAGAGATGTATGCGGAAATGAACAG 1635
Db 412 AsnAlaArgValArgLeuTyrLysProMetValGluGluMetTyrLeuGluThrLys 431

```

RESULT 3

US-10-225-066A-336

; Sequence 336, Application US/10225066A

```

; Publication No. US20030226173A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; INVENTOR: RATCLIFFE, Oliver
; APPLICANT: RIECHMANN, Jose Luis
; APPLICANT: ADAM, Luc J
; APPLICANT: DUBELL, Arnold T
; APPLICANT: HEARD, Jacqueline E
; APPLICANT: PILGRIM, Marsha L
; APPLICANT: JIANG, Cai-Zhong
; APPLICANT: REUBER, T. Lynne
; APPLICANT: CREELMAN, Robert A
; APPLICANT: PINEDA, Omaira
; APPLICANT: YU, Guo-Liang
; APPLICANT: BROUN, Pierre E
; TITLE OF INVENTION: Yield-Related Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MEI0036-2 US
; CURRENT APPLICATION NUMBER: US/10/225,066A
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 09/837,444
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 1122
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 336
; LENGTH: 680
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-10-225-066A-336

Alignment Scores:
Pred. No.: 2,61e-37 Length: 680
Score: 472.00 Matches: 152
Percent Similarity: 42.01% Conservative: 74
Best Local Similarity: 28.25% Mismatches: 193
Query Match: 14.37% Indels: 119
DB: 12 Gaps: 13

US-09-423-575-1 (1-1886) x US-10-225-066A-336 (1-680)
QY 283 ATAAAGTTTCATGAAACCCCAATGGACACACAAACAAACACACTTTTAGTCTCTG 342
Db 32 ValGlnTyrThrGlnGlnAspAspSerAsnAsnAsnAsnAsnSerAsnAsn 51
QY 343 GAT-----AATGTCATGACTACCAAAATCCTTCTCATGGATTTTATACCTCA 393
Db 52 AsnAsnAsnThrAsnThrAsnThrAsnAsnAsnSerSerPheValPheLeuAspSer 71
QY 394 AGAGAAGATTCACTTCATCTCAACAATGTTCCATGGAAATACCATCAGATCAGATCCT 453
Db 72 HisAlaPro-GlnProAsnAlaSerGlnGlnPheValGlyLeuProLeu----- 87
QY 454 CTACAAATGGGTGGCTTTGATATTTTCAATTCATGCTGACTACAAATATCTATCTCT 513
Db 88 -----SerGlyHis-- 90
QY 514 TCTCCAGGCTATCGATGTTCAAGATAACCGAATGTTGAGTTCATGGCTCTCTCTCTCT 573
Db 91 GluAlaAlaSerIleThrAlaAlaAspAsnIleSerValLeuHisGlyTyrProProArg 110
QY 574 CATCCTCCTCAGTTCATCTCTTTGGATCATTAAAGACACTGATGATCTCTCAACAC 633
Db 111 ValGlnTyrSerLeuTyrGly-----SerHisGlnValAspProThrHisGln 126
QY 634 ATGTTGGGTTTGAACCAATAGTAGTTTCAGGATTTTCAGGTGATGTTGGTCCAGT 693

```


Dbb 32 ValGlnTyrThrGlnGlnAspSerAsnAsnAsnAsnAsnSerAsnAsnSerAsn 51
QY 343 GAT-----AATGTCATGACTAACCACAAATCTCTTCTCATGAGATTTTATACITTC 393
Db 52 AsnAsnAsnThrAsnThrAsnThrAsnAsnAsnAsnSerSerPheValPheLeuAspSer 71
QY 394 AGAGAAGATTCAACTTCATCTCAACAATGCTTCCATGCAATACCATCAGATCAGATCCT 453
Db 72 HisAlaPro-GlnProAsnAlaSerGlnGlnPheValGlyLeuProLeu----- 87
QY 454 CTACAAATGGGTGGCTTTGATATTTTCAATTCATGCTGACTAACAAATATCATCTCT 513
Db 88 -----SerGlyHis-- 90
QY 514 TCTCCAGGCTCATGATGTTCAAGATAACCCCAATGTTGAGTTCATGCTCTCTCTCTCT 573
Db 91 GluAlaAlaSerIleThrAlaAlaAspAsnIleSerValLeuHisGlyTyrProProArg 110
QY 574 CATCTCTCCCTCCATCTCTTGGATCATTTAAGACACTATGATGATTCCTCAACAAC 633
Db 111 ValGlnTyrSerLeuTyrGly-----SerHisGlnValAspProThrHisGln 126
QY 634 ATGTGGGGTTTGAGCAAAATAGTGAGTTTCAGGCATTTTCAGGTGTAGTTGGTCCAAAGT 693
Db 127 GlnAlaAlaCysGluThrProArgAlaGlnGlnGlyLeuSerLeuThrLeuSerSerGln 146
QY 694 GAACCAATGATGCTACATTCGGTGAAGAAGATTTCCCGTTTCTTAATTCGAATAAAGA 753
Db 147 GlnGlnGlnGlnGlnHisGlnHisGlnProIleHisValGlyPheGlySer 166
QY 754 AACAATGAGCTTTCATGAGTTCATGCTGATGCTGATGTTTCTGATGATGCTCGAGATAAGT 813
Db 167 GlyHisGlyGluAspIleArgValGlySer----- 176
QY 814 CTTTGTGAGCTACAGATTACCTCAGAGCAGGCTTCTTGACAGCAGCAAGACATTTCT 873
Db 177 -----GlySerThr 179
QY 874 AATAACGTTGTACTCAAGGTTTCTCAACTATATATTTGGCTCAAAATACCTTCACTCT 933
Db 180 GlySerGlyValThrAsnGlyIleAlaAsnLeuVal---SerSerLysTyrLeuLysAla 198
QY 934 GTTCAAGAAATACTATCTCAATTCGCGCATCTGCTGATGATGATGATGATGATGATGAT 212
Db 199 AlaGlnGluLeuLeuAspGluValValAlaAsnAlaAspSerAsp----- 212
QY 994 GAGTCAGGAGCTGAGTTCACCTTTTACTTCACTGCTTTTGAGATATA---ACTGAGTTT 1050
Db 213 AspMetAsnAlaLysSerGlnLeuPheSerSerLysLysGlySerCysGlyAsnAspLys 232
QY 1051 CTTGATGTCGATTTAATAACTCGGAGCGGGTTCGGATCT----- 1092
Db 233 ProValGlyGluSerSerAlaGlyAlaGlyGlyGlyGlySerGlyGlyValaGluAla 252
QY 1093 -----ACATTTCAAGGAGAGCATTAGAAGCAAGAA 1125
Db 253 AlaGlyLysArgProValGluLeuGlyThrAlaGluArgGlnGluIleGlnMetLysLys 272
QY 1126 ACCCATCTCTTGGATCTTCTCAATGTTGATGATGATGATGATGATGATGATGATGATGAT 1185
Db 273 AlaLysLeuSerAsnMetLeuHisGluValGluGlnArgTyrArgGlnTyrHisGlnGln 292
QY 1186 ATTCATACGGTTATATCAGCTTCCATGCTGCAACCGAGTTA---GATCCAGATTACAC 1242
Db 293 MetGlnMetValIleSerSerPheGluGlnAlaAlaGlyIleGlySerAlaLysSerTyr 312
QY 1243 ACCCGGTTTGGCTCCAAACCGTTTCTTCTTATACAAAGACCTCGAGAGAGATCTGC 1302
Db 313 ThrSerLeuAlaLeuLysThrIleSerArgGlnPheArgCysLeuLysGluAlaIleAla 332
QY 1303 AAGAAGATAATCTCTATGCGATCTCTATTTGGAGAGAGGCAACACAGACATCAAGAAC 1362
Db 333 -----GlyGlnIleLysAlaAlaAsnLysSerLeuGlyGluGluAsp 346

QY 1363 TCTATG-----TTCACCCAGCATTCGCTT 1386
Db 347 SerValSerGlyValGlyArgPheGluGlySerArgLeuLysPheValAspHisHisLeu 366
QY 1387 CTTCAAGCAG-----CTGAACGAAAGAACCATCAGATTGG 1422
Db 367 ArgGlnGlnArgAlaLeuGlnGlnLeuGlyMetIleGlnHisProSerAsnAlaIleP 366
QY 1423 AGACCTCAACGAGGTTTGGCTGAGAAATCTGTTTCGGTTCAGGAATTCGATGTTCCAA 1482
Db 387 ArgProGlnArgGlyLeuProGluArgAlaValSerValLeuArgAlaIlePheGlu 406
QY 1483 AACTTCCTTCACTTACCCGAAAGATTCGAGAAACATCTTCTAGCTATACGAATGGC 1542
Db 407 HisPheLeuHisProTyrProLysAspSerAspLysHisMetLeuAlaLysGlnThrGly 426
QY 1543 TTGACAAAGAGTCAGGTATCAAACTGTTTATATAATGCGCGGTTAGGCTATGCAAGCCG 1602
Db 427 LeuThrArgSerGlnValSerAsnTyrPheIleAsnAlaArgValArgLeuTyrLysPro 446
QY 1603 ATGATAGAAGAGATGATGCGGAAATGCAACAGAGAGCTCAATTAACATCATTCATCAA 1662
Db 447 MetValGluGluMetTyrMetGluMetLysGluGlnAlaLysAsnMetGlySerMet 466
QY 1663 CCCAACCGACCA-----ACTCTTCGAATG 1686
Db 467 GluLysThrProLeuAspGlnSerAsnGluAspSerAlaSerLysSerThrSerAsnGln 486
QY 1687 CCAAAATCTGTTATGATGAGCCCAAGCAATGCATAAATAAGACAAACAAT 1734
Db 487 GluLysSerProMetAlaAspThrAsnTyrHisMetAsnProAsnHis 502

RESULT 5
US-10-425-114-56821
; Sequence 56821, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 56821
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE: Zea mays
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17364H11_FLI.pep
US-10-425-114-56821

Alignment Scores:
Pred. No.: 6,27e-37 Length: 475
Score: 467.50 Matches: 144
Percent Similarity: 46.67% Conservative: 80
Best Local Similarity: 30.00% Mismatches: 175
Query Match: 14.24% Indels: 81
DB: 12 Gaps: 14

US-09-423-575-1 (1-1886) x US-10-425-114-56821 (1-475)

QY 292 CATGAAGAACCATGAGACACACACACACAACTTTTCTGATATGTC 351
Db 2 HisGlnSerAspSerAspAsnAsnSerHisTyrLeuProSerSerLeuHisArgSer 21
QY 352 ATGACTAACCAAAATCCTCTTCTCTGATTTTATACCTTCAAGAGAGATTCACATTC 411

```

Db      22 ValSerTrpArgArgLeuValValLeu-----LysThrMetAlaThrPhe 36
QY      412 TTCTCAACAGCTTCCATGGAAATACCATCAGATCAGATCCTCTACAAATGGGTGGCTTT 471
Db      37 PheSerThr-----SerThrAsnGlnArgAspLeuAlaGlyGlyGlyGlyGlyGly 51
QY      472 GATATTTTCAATCTCTGCTGACTACAAATACCTATCATCTCTCCACGGTCTCTATCGAT 531
Db      52 -----Asp 52
QY      532 GTTCAAGATAACCGCAATGTGAGTTCATGGCTCCTCCT-----CCTCAT 576
Db      53 MetSerSerPheHisHisTyAlaTySerAspProAlaGlyGlyLeuMetProPhe 72
QY      577 CCTCTCCACTTCATCCTTGGAT---CATTTAGACACTATGATGATTCCTCAACCAAC 633
Db      73 ProAlaThrIleValSerSerGluGlyHisValAlaHisGlyGlyAspGly----- 90
QY      634 ATGTGGGTTTTGAAGCAATAGTAGTTTCAGGCAATTTTCAGGTGTAGTTGGTCCAAGT 693
Db      91 -----GlyArgGlyAlaPheValAspValAsnAlaArgAspGlyProThrGlyGlyAla 108
QY      694 GAACCAATGATGCTACATTCGGTGAAGAAATTTCCCGTTCTTAATTTGCAATAAA--- 750
Db      109 GluMetGlyLeuGlnThr-----GlnLeuLeuMetAlaAsnAlaAla 122
QY      751 -----AGAAACAATGAGCTTTTCATTTGAGTCTTTCATCAGAT----- 786
Db      123 SerAlaValGlnHisGlnGlyLeuSerLeuSerLeuGlyThrGlnGlyValProValSer 142
QY      787 ---GTTTCTGAGTAAGTCTGGAGATAAGTCTTTGTGAGTCAAGATAGCCTCAGAG 843
Db      143 LeuTyGlnTyArgGlnAlaGlyMetAlaAlaAlaPheLeuSerProAsn 162
QY      844 CAAGCTTCTTCAGACGAAAGACATTTCTAATACGTTGTACTCAAGGTTTCTCTCAA 903
Db      163 GlnAlaSerValAlaAlaThrArgSerAlaGlnSerIleTyValGln----- 178
QY      904 CTATATTGGCTCAAAATACCTTCACCTCTGTTCAAGAAATACATCTCAATTCCGCC--- 960
Db      179 -----AsnSerArgTyLeuLysAlaAlaArgGluLeuLeuAspGluValValAsn 195
QY      961 -----GCATACCTCGCTCATTTATTCATCTCAGGACCGAGTCAGGAGCT 1005
Db      196 ValGlnAspAlaIleLysArgLysGlyAspLysSerGlnGlnGlyLysAspSerGlyGly 215
QY      1006 GCTAGTTCAGCCTTTTACTTCACTGTTTGAATATACTGAGTTTCTTTGATGGTGAT--- 1062
Db      216 GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 235
QY      1063 TCTAATACTCGGAGCGGGTTTCGATCTCATTTCAAGAGAGCAATAGAACGCAAG 1122
Db      236 GluGlyAsnSerSerAlaProGluLeuSerProSerGluArgGlnAspLeuGlnAsnLys 255
QY      1123 AAAACCCATCTCTGGATCTTCTTCAATGGTGGATGATCATATAGTCATTCGCTAGAT 1182
Db      256 ValSerAlaLeuMetAlaLeuLeuAspGlnValAspArgLysTyArgHisTyHisHis 275
QY      1183 GAGATTCTACGGTTATATCAGCGTTCCATGCTGCAACCGAGTTAGAT---CCACAGTTA 1239
Db      276 GlnMetGlnMetValMetSerSerPheAspAlaValAlaGlyAlaGlyAlaAlaArgPro 295
QY      1240 CACACCGGTTTGGCTTCAACCGCTTCTTCTTATACAAAGACCTCGAGAGAGATC 1299
Db      296 TyThrAlaLeuAlaLeuGlnThrIleSerArgHisPheArgSerLeuArgAspAlaVal 315
QY      1300 TGCAAGAAGATAATCTCTATGGGATCTCTATTTGAGAGAGCAAGCAAGACTCAAGAA 1359
Db      316 GlyAlaGlnValGlnSerLeuArgSerLeuGlyGlyGlyGlyGlyGlyGlyGlyGly 335
QY      1360 ACCTCTATGTTCCACGACATTCCTCTTTCAGCAGCTGAACGAAACAAC----- 1410

```

```

Db      336 GlyGlyLeuProArgLeuArgTyIleAspGlnGlnLeuArgGlnGlnArgAlaMetGln 355
QY      1411 -----CATCAGATTTCGAGACCTCAACGAGGTTTGCCTGAG 1446
Db      356 GlnPheGlyMetMetGlnGlnProGlnHisAlaIlePArgProGlnArgGlyLeuProGlu 375
QY      1447 AAATCTGTTTCGGTTCCTACGGAATTCGATGTTCCAAAACCTTCCTTACCCCTTACCCGAAA 1506
Db      376 SerAlaValSerValLeuArgAlaIlePLeuPheGluHisPheLeuHisProTyProLys 395
QY      1507 GATTCGGAGAAACATCTTCTAGCTATACGAGTGCCTTGACAAAGTCAAGTATCAAC 1566
Db      396 AspSerGluLysLeuMetLeuAlaArgGlnThrGlyLeuSerArgGlyGlnValSerAsn 415
QY      1567 TGTGTTTATAATCGCGGGTTCAGGCTATGGAAGCCGATGATAGAAGAGATGTATCGGAA 1626
Db      416 TrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluGluMetTyArgGlu 435

RESULT 6
US-10-425-114-68805
; Sequence 68805, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yinhua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68805
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLM017199B09_FLI.pep
US-10-425-114-68805

Alignment Scores:
Pred. No.: 6,4e-37 Length: 499
Score: 467.50 Matches: 117
Percent Similarity: 54.09% Conservative: 55
Best Local Similarity: 36.79% Mismatches: 103
Query Match: 14.24% Indels: 43
DB: 12 Gaps: 10

US-09-423-575-1 (1-1886) x US-10-425-114-68805 (1-499)
QY      763 CTTTCATTGAGTCTTCATCAGATGTTTCTGATGAATGCTCGAGATAGTCTTTGTGCA 822
Db      79 LeuSerLeuSer-----SerArgGluAlaProProValThrValAlaAla 93
QY      823 GCTACAAGATTAGCCTCAGAGCAAGCTTCTTCAGC-----AGCAAGACATTTCTAAT 876
Db      94 Val-----AlaAlaCysThrAspGluGlyLysTyLeuLysVal 106
QY      877 AACGTTGTTTACTCAAGGTTTCTCTCAACTATATTTTGGCTCAAAATACCTTCCTGTT 936
Db      107 SerAlaThrProGluGly-----GlnMetValMetSerSerLysTyLeuLysAlaAla 124
QY      937 CAAGAAATACTATCTCATTTCCCGCATCTCGCTCGATTATTCATCTCGAGGAACCGAG 996
Db      125 GlnGluLeuLeuAspGluValValSerValSerLysGlyValGluAspAlaLysThrAla 144
QY      997 TCAGGAGCTGCTAGTTTCAGCCTTTTACTTCA---CGTTTGGAGATATTAAGTCTTCT 1053
Db      145 AlaAlaAlaAlaThrLysSerLeuAlaAlaValLysLysLysLysLysLysLysLysLys 164

```


QY 1054 GATGGT-----GATTCTTAATACTCGAGCGGGTTTCGGA----- 1089
Db 165 SerGlyGlyThrAspAspGlyAlaGlyAlaLysSerGlyGlyAlaAlaProGlu 184
QY 1090 ---TCTACATTCGAAGAGACATTAGAACGCAAGAAACCCATCTCTTGATCTTCT 1146
Db 185 MetSerThrAlaGluArgGlnMetGlnMetLysSerLysLeuLeuAsnMetLeu 204
QY 1147 CAATGGTGGATGATCATATAGTTCGCTGATGATGATTCATACCGTTTATATACAG 1206
Db 205 AspGluValGluGlnArgTyrArgGlnTyrHisGlyGlnMetGlnAlaValSerSer 224
QY 1207 TTCCATGCTGCACCGAGTTA---GATCCACAGTTACACACCGGTTGCTCCCAAC 1263
Db 225 PheGluAlaAlaGlyAlaGlySerAlaArgThrTyrThrAlaLeuAlaLeuArgThr 244
QY 1264 GTTTCCTTCTTATACAAAGACCTGACAGAGAGAACTGCAAGAGATTAATCTCTATGGA 1323
Db 245 IleSerArgGlnPheArgCysLeuArgAspAlaIleAlaSerGlnValArgAlaAlaSer 264
QY 1324 TGTGATTGGAGAGGCAAGACAGACTCAAGAAACCTCT----- 1365
Db 265 ArgAlaLeu-----GlyGluAspAlaAspAlaAlaValAlaGlyGlyArgThrValGly 282
QY 1366 -----ATGTTCCACGACATTGCTTCTTCAGCAGCTGAAA 1401
Db 283 SerArgLeuArgTyrIleAspHisGlnLeuArgGlnArgAlaLeuGlnLeuGly 302
QY 1402 CGAAGAACCATCAGATTGGAGACCTCAACGAGGTTGCTCGAAGAAATCTGTTTCGTT 1461
Db 303 MetMetGlnGlyGlyAlaTrpArgProGlnArgGlyLeuProGluArgSerValSer 322
QY 1462 CTACGGAATGATGTTCCAAACTCTTCACCTTACCCGAAAGATTCGAGAGAAACAT 1521
Db 323 LeuArgAlaTrpLeuPheGlnHisPheLeuHisProTyrProLysAspSerAsp 342
QY 1522 CTTCAGTATACGAAGTGGCTTGACAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1581
Db 343 MetLeuAlaLysGlnThrGlyLeuThrArgSerGlnValSerAsnTrpPheIleAsnAla 362
QY 1582 CGGTTAGCTATGAGCCGATGATAGAGAGATGATGCGGAATGACAG 1635
Db 363 ArgValArgLeuTrpLysProMetValGluGluMetTyrLeuGluGluThrLys 380

RESULT 7
US-10-437-963-122156
; Sequence 122156, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122156
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_2510C.1.p
US-10-437-963-122156

Alignment Scores:
Pred. No.: 6.89e-37 Length: 594

Score: 467.50 Matches: 123
Percent Similarity: 45.90% Conservative: 56
Best Local Similarity: 31.54% Mismatches: 98
Query Match: 14.24% Indels: 113
DB: 16 Caps: 10

US-09-423-575-1 (1-1886) x US-10-437-963-122156 (1-594)

QY 757 AATGAGTTTCATGAGTCTTGTCATCAGATGTTTCT----- 792
Db 182 SerGluLeuSerLeuSerLysSerLysSerSerAspSerMetLeuAsnAlaGly 201
QY 793 ---GATCAATCTCGGAGATAAGCTTTTGTGACGTACAGATTAGCTCCTCAGAGCAAGCT 849
Db 202 GlyAspGlnCysSerSer-----AlaAlaSerArgSerGlyLeuThrGln--- 216
QY 850 TCTTGCAGACCAAGACATTTCTTAATACGTTGTTACTCAA----- 891
Db 217 ---MetSerArgValValValGluProGluProGluProLeuVal 230
QY 892 -----GTTTCTCTCAACTTATATTTGGCTCAAAATACCTTCACTCT 933
Db 231 ProTyrTyrProAlaAlaAsnPheAlaValValAlaArgSerArgTyrAlaAlaVal 250
QY 934 GTTCAAGAAATATCTATCTCATTTCT----- 957
Db 251 AlaGlnGlnValLeuAsnAspAlaValGlyCysValLeuGlyValAlaAspAlaAla 270
QY 958 -----CCGCATCTACTCGCTCGATTAT---TCATCTCGAGAACCGAGTCAGAGCTGCT 1008
Db 271 AlaAspSerAlaSerGlyValAspSerGlySerSerArgProSerSerCysSerValAla 290
QY 1009 AGTTCAGCCTTTACTTCACGTTTTGAGAATATACTGAGTTTCTTGTGATGTTCTAAT 1068
Db 291 GlyGlyAlaProSerSerAlaValSerSerAsnAsnGlnLeuIleAlaSerSerGlyGlu 310
QY 1069 AACTCGAGGCGGTTTCGGATCTACATTTCAAAGAGAGCATTAGAAGCAAGAAAC 1128
Db 311 HisThrHisGlyGlyGlyAspAlaSerAlaGlnArg-----LeuArgSer 325
QY 1129 CATCTCTGATCTCTTCAAAATGGTG----- 1155
Db 326 GluLeuLeuThrMetLeuGlnLeuValSerSerProPheSerThrArgLeuProArgHis 345
QY 1155 ----- 1155
Db 346 GlnPheLysGlnIleSerTyrTyrHisPheGluSerAsnLeuLysProSerPheAspAsp 365
QY 1156 -----GATCATCGATATAGTCTATTCCTGCTAGATGATTCAT 1191
Db 366 AspThrAsnSerAsnAlaGlnMetAspGlnLysTyrAsnGlnCysLeuAspGluIleGln 385
QY 1192 ACGTTTATATCAGCGTTCCATGCTGCAACCGAGTTA----- 1227
Db 386 SerThrThrAlaArgPheAsnThrLeuThrHisAlaThrAlaArgAlaAlaGlyMetSer 405
QY 1228 GATCCACAGTTACACCCCGTTTGCCTCAACCGTTTCTTCTTATACAGAACCTG 1287
Db 406 SerSerSerIleCysAlaProPheAlaHisArgAlaValSerAlaMetTyrHisGlyLeu 425
QY 1288 AGACAGAGAAATCTGCAAGAGATAATCTCTATGGATCTGTTATGGAGAGA----- 1338
Db 426 ArgArgGlnAlaGlyGluIleMetSerAlaAlaAlaAlaAlaGlyArgProCysArg 445
QY 1339 -----GGCAAGACAGACTCAAGAAACCTCT 1365
Db 446 GlyGlyGluSerSerSerAlaValThrGlyGlyGluArgGluArgSerTrpGluSerAla 465
QY 1366 ATGTTCCACAGCATTCCTTCTCAGCAGCTGAACGAAAGAACCATCAGATTGAGAGA 1425
Db 466 PheIleGlnLysHisTrpAlaValGlnGlnLeuArgGlyGluGlnGlnCysTrpArg 485
QY 1426 CCTCAACGAGGTTTGCCTGAGAAATCTGTTTCGGTTCTACGGAATGATGTTCCAAAC 1485

Db 486 ProGlnArgGlyLeuProGluLysSerValAlaValLeuLysAlaTrpMetPheGluAsn 505
 QY 1486 TTCCTTACCTTACCGAAGATTCGAGAACATCTTCTAGCTATACGAGTGGCTTG 1545
 Db 506 PheLeuArgProTyrProLysAspSerGluLysGluMetLeuAlaAlaArgSerGlyLeu 525
 QY 1546 ACAAGAGTCAGTATCAACCTGGTTTATAAATGCGCGGTAGGCTATGGAAGCCGATG 1605
 Db 526 SerArgAsnGlnValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysPromet 545
 QY 1606 ATAGAAGAGATGATCGCGAATGAAACAAAG 1635
 Db 546 IleGluGluMetCysGluGluLeuLysArg 555

RESULT 8

US-10-437-963-110193
 ; Sequence 110193, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 110193
 ; LENGTH: 618
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_1427C.1.pbp
 US-10-437-963-110193

Alignment Scores:
 Pred. No.: 1248-36 Length: 618
 Score: 485.00 Matches: 141
 Percent Similarity: 46.53% Conservative: 60
 Best Local Similarity: 32.64% Mismatches: 119
 Query Match: 14.16% Indels: 112
 DB: 16 Gaps: 16

US-09-423-575-1 (1-1886) x US-10-437-963-110193 (1-618)

QY 583 CCATCTATCTCTTGGATCATTTTAAAGACATATGATGATCTCTCAAAACAAATGTGGGT 642
 Db 147 ProLeuHisGlyAlaTyrTyrLeuSerGlyPhe-----SerSerGlyAlaGlyAlaGly 164
 QY 643 TTGGAAGCAAAATGAGTGGTTCAGGCATTTTCAGGTGTAGTTGGTCCAAAGTGAACCAATG 702
 Db 165 Phe-----LeuSerProPheAlaAlaSer-----ValAlaAlaAlaProAla 180
 QY 703 ATGCTTACATCCGTGAAGAGATTTCCCGTTCTTAATTTCCAAATAAAGAAACAATGAG 762
 Db 181 AlaSer-----Glu 183
 QY 763 CTTTCATTGAGTCTTGCA-----TCAGATGTTCTGATCAATGC 801
 Db 184 LeuSerLeuArgLeuGlyAlaThrLysCysSerSerProSerSerMetAlaAsnAlaSer 203
 QY 802 TCGAGATAAAGTCTTTGTGACAGCTACAAGATTAGCTCA-----GAG 843
 Db 204 SerGluValSerCysSerGlyLeuThrHisValSerSerGlyGlyLeuGlyTyrHis 223
 QY 844 CAACTTCTTTCAGCAGCAAGACATTTCTTAATAACGTGTGTACTCAAGGT----- 894

Db 224 GluAlaAlaAlaGlyAlaGlyAlaAlaLeuPheHisProThrHisGlyAspAspAla 243
 QY 895 -----TTC 897
 Db 244 AlaAlaAlaAlaGlyGluLeuArgGlnAlaTyrHisSerArgAlaProProHisPhe 263
 QY 898 TCTCACTTATATTGGCTCAAAATACCTT---CACTCTGTTCAAGAAATATCTATCTCAT 954
 Db 264 SerGlnValValSerArgSerAlaValLeuAlaHisValAlaGlnGluLeuLeuAsnGly 283
 QY 955 TTGCGCGCATACTCGCTCGAT-----TATTCTCTCGAGAAACCGAGTCAGGA 1002
 Db 284 PheValAlaCysLeuLeuGlnAspValAlaAlaAspAlaAlaSerGlyValAspGlyGly 303
 QY 1003 GCTGCTAGTTCAGCCTTTACTTCCAGCTTTTGAGAATATACTGAGTTTCTGATGCTGAT 1062
 Db 304 GluAlaSerGlnAla-----LeuLeuValGlyLeuPheGlyGlu 316
 QY 1063 TCTAATAAATCGGAGCGGGTTCGGATCTTCAATTTCAAAGGAGA----- 1107
 Db 317 AspHisAspGlyAlaAspGlyGlyCysLeuThrArgGluArgArgCysGluValGlyGly 336
 QY 1108 -----GCATTAGAACAAAGAAACC----- 1128
 Db 337 GlyGlyAlaAlaSerGluGluAlaThrProThrGlyGluPheLeuAlaGlnPheLeuVal 356
 QY 1129 ---CATCTCTTGGATCTTCTT-----CAATGTGGTATGAT 1161
 Db 357 GluHisLeuArgAsnArgPheLeuValArgThrAsnGlyAspGluMetGlnValAspGlu 376
 QY 1162 CGATATAGTCATTCGTAGATGATCATACAGTTATATACGGTTATATACGGTTCATCGTCAACC 1221
 Db 377 LysCysAsnGlnCysValGluGluMetGlnSerThrAlaAlaArgPheAsnSerMetVal 396
 QY 1222 GAGTTA-----GATCCACAGTTACACACCGGTTTCCCTCCAAACCGTTCC 1269
 Db 397 ArgSerThrGlyGlyGlyGlyLeuThrAlaAlaPheAlaGlyArgAlaValAla 416
 QY 1270 TTCTTATACAAGAACCTGAGAGAGAAATCTCAAGAGATATATC----- 1314
 Db 417 AlaAlaTyrArgArgValArgValMetGlyGlnLeuValAlaAlaAlaThrAla 436
 QY 1315 -----TCTATGGATCTGTATTGGAGAGGCAAGACAGACTCAAGAA 1359
 Db 437 ArgSerSerSerAlaAlaAlaAlaLeuGluGlu---LysGluArgSerTrpGlu 455
 QY 1360 ACCTCTATGTTCCACCGATTCCTTCTTACAGCAGTGAACAGAAACCATCAGATT 1419
 Db 456 SerSerPheIleGlnLysHisTrpAlaMetGlnGlnLeuArgArgGlyAspGlnGlnSer 475
 QY 1420 TGAGACCTCAAGAGGTTTGGCTGAGAAATCTGTTTCGGTCTACGGAATGGATGTTTC 1479
 Db 476 TrpArgProGlnArgGlyLeuProGluLysSerValAlaValLeuLysAlaTrpMetPhe 495
 QY 1480 CAAATCTTCTTCCACCTTACCCGAAAGATTCGGAGAACATCTTCTAGCTATACGAGT 1539
 Db 496 GluAsnPheLeuArgProTyrProLysAspSerGluLysAspMetLeuAlaAlaArgSer 515
 QY 1540 GGTTCGACAAAGATCAGGTATCAAACTGGTTTATAAATGCGCGGTAGGCTATGGAAG 1599
 Db 516 GlyLeuSerArgSerGlnValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLys 535
 QY 1600 CCGATCATAGAAGATGATCGGAAATGAACAAG 1635
 Db 536 ProMetIleGluAspMetTyrGluGluLeuLysLys 547

RESULT 9

US-10-437-963-162637
 ; Sequence 162637, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 162637
 ; LENGTH: 678
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_6170C.1.p
 ; US-10-437-963-162637

Alignment Scores:
 Pred. No.: 1,59e-35 Length: 678
 Score: 454.00 Matches: 137
 Percent Similarity: 46.93% Conservativity: 62
 Best Local Similarity: 32.31% Mismatches: 149
 Query Match: 13.82% Indels: 76
 DB: 16 Gaps: 14

US-09-423-575-1 (1-1886) x US-10-437-963-162637 (1-678)

QY 481 AATTCTATCTGACTACAAATCTATCATCT-----513
 DB 48 AsnMetMetLeuAsnSerAlaValThrSerMetThrProValSerPheGlyHisGln 67
 QY 514 ---TCTCCACGGTCTATCATGTTCAAGATAACCGCAATGTTGAGTTCATGCTCCTCCT 570
 DB 68 ProSerProSerSerSerSerAlaAlaGlnHis-----PheValGlyLeuPro 83
 QY 571 CTTCTATCTCTCCACTCATCTCTTGGATCATTTAAGACACTATGATGATCTCTCAAC 630
 DB 84 LeuGlnAlaProProAla-----SerGlyTyr 92
 QY 631 AACATGCTGGGTTTTCGAACAAATAGTGTGATTTTCAGGCATTTCAGTGTAGTGTGCTCA 690
 DB 93 AsnLeuThrProAlaAlaThrGly-----AlaGlyAspMetSerPro 108
 QY 691 AGTGAACCAATGATGCTTACATTCGGTGAAGAAGATTTCCCGTTCTTAATTCGAATAAA 750
 DB 109 ProThrPro-----GlnHis 113
 QY 751 AGAAACAATGAGCTTTTCATTGAGTCTTCATCAGATGTTTCTGATCAATGCTCGGAGATA 810
 DB 114 GlnHisGlnAlaHisGlyGlyAlaAlaGlyValSerAlaValLeuSerLeuSer 133
 QY 811 AGTCTTTGTGCGACT-----ACAAGATTAGCTCAGACCAAGCTTTCTTCGACGACG 861
 DB 134 SerArgGluAlaAlaProProValThrValAlaAlaValAlaAlaGlyAspGluGly 153
 QY 862 AAGACATTTCTTAATACGTTGTTACTCAAGT-----TCTCTCACTTATATT 912
 DB 154 LysTyrLeu-----GlnAlaValAlaGlnGlyAlaAlaSerHisGlyGlnMetValMet 171
 QY 913 GGTCTAAATACCTTCACCTCTGTTCAGAAATATCTATCTCAATTCGCCGCATCTCGCTC 972
 DB 172 SerSerLysTyrLeuLysAlaAlaGlnGluLeuLeuAspGluValValSerValSerLys 191
 QY 973 GATTATTTCATCTCGAGAACCGAGTCAGGAGCTGCTAGTTCAGCTTTTACTTCACGTTTT 1032
 DB 192 GlyValAspValLysAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 211
 QY 1033 GAGAATAATACTGAGTTCTTGTGCT-----CATTTCTAATACTCGGAGCG 1080

DB 212 GluAsp---SerGluGlyValSerGlyGlyGlyThrGluAspGlyGlyGlyAlaLysSer 230
 QY 1081 GGTTCGGA-----TCTACATTTCAAGGAGAGCATTAGAACAAAG 1122
 DB 231 GlyGlyAlaProProProGluMetSerThrAlaGluArgGlnGluLeuGlnMetLys 250
 QY 1123 AAAACCCATCTCTGGATCTCTTCAATGCTGATGATCGATATAGTCACTCGGTAGAT 1182
 DB 251 LysGlyLysLeuLeuAsnMetLeuAspGluValGluGlnArgTyrArgGlnTyrHisGln 270
 QY 1183 GAGATTCATACGGTTATATCAGCGTTCATGCT---GCAACCGAGTTAGATCCACAGTTA 1239
 DB 271 GlnMetGlnValValValAlaSerPheGluAlaValAlaGlyGlyGlySerAlaArgThr 290
 QY 1240 CACACCGGTTTGGCTCCAAACCGTTTCTTTATACAAGAACCTGAGAGAGAGATC 1299
 DB 291 TyrThrAlaLeuAlaLeuArgThrIleSerArgGlnPheArgCysLeuArgAspAlaIle 310
 QY 1300 TGCAGAGATAATCTCTATGGGATCTGTATTG-----GAGAGAGGC 1341
 DB 311 AlaGlyGlnValArgAlaAlaSerArgAlaLeuGlyGluAlaValAspAlaAspGlyGly 330
 QY 1342 AAAGACAAGACTCAAGAAACCTCTATG-----TTCACCAAGCATTCG 1383
 DB 331 CysGlyArgThrValGlySerArgLeuArgTyrIleAspHisGlnLeuArgGlnArg 350
 QY 1384 CTTCTCAGCAGCTGAAACGAAAGACCATCAGATTTGGAGACTCAACGAGTTTCCT 1443
 DB 351 AlaLeuGlnGlnLeuGlyMetGlnSerSerAlaTrpArgProGlnArgGlyLeuPro 370
 QY 1444 GAGAAATCTGTTTCGGTTCACGAAATGGATGTTCCAAACTTCCTTACCCCTTACCCG 1503
 DB 371 GluArgSerValSerIleLeuArgAlaTrpLeuPheGluHisPheLeuHisProTyrPro 390
 QY 1504 AAGATTCGAGAGAAACATCTTCTAGCTATACGATGCTGCTGACAAAGTCAAGTATCA 1563
 DB 391 LysAspSerAspLysIleMetLeuAlaLysGlnThrGlyLeuThrArgSerGlnValSer 410
 QY 1564 AACTGTTTATAATCGCGGTTAGCTATGGAAGCCGATGATAGAGAGATGTATCGG 1623
 DB 411 AsnTrpPheLeuAsnAlaArgValArgLeuTrpLysProMetValGluGluMetTyrLeu 430
 QY 1624 GAAATGAACAAG 1635
 DB 431 GluGluThrLys 434

RESULT 10
 US-10-424-599-180416
 ; Sequence 180416, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 180416
 ; LENGTH: 538
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_133930C.1.p
 ; US-10-424-599-180416

Alignment Scores:
 Pred. No.: 1.77e-34 Length: 538
 Score: 443.00 Matches: 133
 Percent Similarity: 46.67% Conservativity: 77

```

Best Local Similarity: 29.56% Mismatches: 152
Query Match: 13.49% Indels: 88
DB: 12 Gaps: 17

US-09-423-575-1 (1-1886) x US-10-424-599-180416 (1-538)

QY 586 CTCATCCCTTGGATCATTTAGACACATGATGATTC---TCAACAACATCGTGCGGT 642
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 2 LeuAsnThrLeuHisLeuLeuLeuHisHisAspGlnArgGileAArgSerHisLeuTrp-- 20
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 643 TTTGAACAAATAGTAGTTCAGGCATTTCAGGTGTAGTTGGTCCAAGTGAA----- 696
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 21 -----PheGluLeuThrProSerValHisProProSerSerThrThr 34
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 697 -----CCAATGANG-----TTCATCATTGGTGAAGAA 723
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 35 ThrValProGlnIleThrLysThrGlnArgAlaThrAsnSerAlaLeuGlyGln-- 53
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 724 CAITTCOCGTTCTTAATTCGAATAAAGAACAATGAGCTTCATTGAGTCCTTGCATCA 783
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 54 -----ProCysLeuLeuSerLeuAlaLeuSerProProLeuAlaTyrrSerLeuLeuArg 71
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 784 GATGTTTCTGTGAATGCTCGGAGATAAGTCTT-----TGTGCAGCTACAAGATTAGCC 837
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 72 GluPheGlySerAlaTrpAlaArgHisProValTrpGlyValProValAlaArgLeuAla 91
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 838 TCAGACGACACTCTTCGACGACGACAAACATTTCTAATAACGTTGTTACTCAAGTTTC 897
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 92 GlnGlnMetLeuSerGlnAspGlyTyrrValSerSerAsnSerLysValAlaSerValTy 111
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 898 TCTCAA-----CTTATATTTGGCTCAAAATACCTTCACTCTGTTCAAGAAATACTA 948
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 112 GlnGlnGlyHisPheLeuValLysAsnSerLysPheLeuValProAlaGlnAspLeuLeu 131
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 949 TCTCATTTGCCGCATCTCGCTCGATTCATCTCTCGAGAACCGAGTCAGAGCTGCT 1008
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 132 AsnGluPheCysSer-----LeuCysAlaLysGlnSerAspLeuGlyLysPro 147
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 1009 AGTTCAGCTTTACTTCACGTTTTGAGAATATACTAGTGTCTTGTGATGTTCTAAT 1068
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 148 ThrLysSerLeuLysGlnTrpGlu-----AspGlnGlu 159
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 1069 AACTCGAGCGGGTTTCGATCTACATTTCAAAGAGAGCATTAGAA----- 1116
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 160 AsnAsnGlyValGlySerSerLysLysHisSerLeuThrSerLeuGluPheValGluLeu 179
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 1117 ---GCAAGAAAACCCTCTCTGATCTCTTCAAATGGTGGATGATCATATAGTCAT 1173
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 180 GlnLysArgLysThrLysLeuLeuSerMetLeuGluValaspArgArgTyrrLysHis 199
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 1174 TGGTAGATGAGATTCATACGGTTATATCACGGTTTCCATGCT---GCAACCGAGTTAGAT 1230
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 200 TyrArgAsnGlnMetLysSerValValSerPheGluAlaValAlaGlyAsnGlyAla 219
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 1231 CCACAGTTACACACCGGTTTGCCTCCAAACCGTTTCTCTTATACAGAACCTGACA 1290
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 220 AlaThrValtyrSerAlaLeuAlaLeuLysAlaMetSerArgHisPheArgCysLeuLys 239
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 1291 GAGAGAAATCTGCAAGAAGATA-----ATC 1314
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 240 AspGlyIleLeuSerGlnIleGlnAlaThrArgLysAlaMetGlyGlyLysAspProVal 259
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 1315 TCTATGGGATCTGTATTGGAGAGCGCAA-----GACAGACT 1353
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 260 AlaProGlyThr-----ArgGlyGluThrProArgLeuLysValIleAspGlnThr 277
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 1354 CAAGAAACCTCTATGTTCCACGACANTGCTCTTTCAGCAGCTGGAACGAAAGAACCAT 1413
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 278 -----LeuArgGlnGlnArgAlaPheGlnGlnMetSerMetMetGluThr 292
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 1414 CAGATTGGAGACCTCAACGAGGTTTGGCTGAGAAATCTGTTTCGGTTCTACGGAATTGG 1473
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 293 HisProTyrrArgProGlnArgGlyLeuProGluArgAlaValSerValLeuArgAlatrp 312
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

Best Local Similarity: 29.56% Mismatches: 152
Query Match: 13.49% Indels: 88
DB: 12 Gaps: 17

US-09-423-575-1 (1-1886) x US-10-424-599-180416 (1-538)

QY	586	CTTCATCCCTTGGATCATTTAGACACATGATGATCC-----TCAACAACATCGTGCGGT	642
D	2	LeuAsnThrLeuHisLeuLeuLeuHisHisAspGlnArgGileAArgSerHisLeuTrp--	20
QY	643	TTTGAACAAATAGTAGTTCAGGCATTTCAGGTGTAGTTGGTCCAAGTGAA-----	696
D	21	-----PheGluLeuThrProSerValHisProProSerSerThrThr	34
QY	697	-----CCAATGANG-----TTCATCATTGGTGAAGAA	723
D	35	ThrValProGlnIleThrLysThrGlnArgAlaThrAsnSerAlaLeuGlyGln---	53
QY	724	GATTTCOCGTTCTTAATTCGAATAAAGAACAATGAGCTTCATTGAGTCCTTGCATCA	783
D	54	-----ProCysLeuLeuSerLeuAlaLeuSerProProLeuAlaTyrsrLeuLeuArg	71
QY	784	GATGTTTCTGTGAATGCTCGGAGATAAGTCTT-----TGTGCAGCTACAAGATTAGCC	837
D	72	GluPheGlySerAlaTrpAlaArgHisProValTrpGlyValProValAlaArgLeuAla	91
QY	838	TCAGACGACACTCTTCGACGACGACAAACATTTCTAATAACGTTGTACTCAAGTTTC	897
D	92	GlnGlnMetLeuSerGlnAspGlyTyrValSerSerAsnSerLysValAlaSerValTyr	111
QY	898	TCTCAA-----CTTATATTTGGCTCAAAATACCTTCACTCTGTTCAAGAAATACTA	948
D	112	GlnGlnGlyHisPheLeuValLysAsnSerLysPheLeuValProAlaGlnAspLeuLeu	131
QY	949	TCTCATTTCCGCGTACTCGCTCGATTCATCTCGAGAACCGAGTCAGAGCTGCT	1008
D	132	AsnGluPheCysSer-----LeuCysAlaLysGlnSerAspLeuGlyLysPro	147
QY	1009	AGTTCAGCTTTACTTCACGTTTTGAGAATATACTAGTGTCTTGTGATGATCTTAAT	1068
D	148	ThrLysSerLeuLysLysGlnTrpGlu-----AspGlnGlu	159
QY	1069	AACTCGAGCGGGTTTCGATCTACATTTCAAAGAGAGCATTAGAA-----	1116
D	160	AsnAsnGlyValGlySerSerLysLysHisSerLeuThrSerLeuGluPheValGluLeu	179
QY	1117	---GCCAAGAAAACCCTCTCTGATCTCTTCAATTTGGTGGATGATCATATAGTCAT	1173
D	180	GlnLysArgLysThrLysLeuLeuSerMetLeuGluValaspArgArgTyrLysHis	199
QY	1174	TGGTAGATGAGATTCATACGGTTATATCACGGTTTCCATGCT---GCCAACGAGTTAGAT	1230
D	200	TyrArgAsnGlnMetLysSerValValSerPheGluAlaValAlaGlyAsnGlyAla	219
QY	1231	CCACAGTTACACACCGGTTTGCCTCCAAACCGTTTCTCTTATACAGAACCTGACA	1290
D	220	AlaThrValTyrSerAlaLeuAlaLeuLysAlaMetSerArgHisPheArgCysLeuLys	239
QY	1291	GAGAGAAATCTGCAAGAAGATA-----ATC	1314
D	240	AspGlyIleLeuSerGlnIleGlnAlaThrArgLysAlaMetGlyGlyLysAspProVal	259
QY	1315	TCTATGGGATCTGTATTGGAGAGCGCAA-----GACAGACT	1353
D	260	AlaProGlyThr-----ArgGlyGluThrProArgLeuLysValIleAspGlnThr	277
QY	1354	CAAGAAACCTCTATGTTCCACGAGCAITTCCTCTTCAGCAGCTGGAACGAAAGAACCAT	1413
D	278	-----LeuArgGlnGlnArgAlaPheGlnGlnMetSerMetMetGluThr	292
QY	1414	CAGATTTGGAGACCTCAACGAGGTTTGCCTGAGAAATCTGTTTCGGTTCTTCGGAATTGG	1473
D	293	HisProTyrArgProGlnArgGlyLeuProGluArgAlaValSerValLeuArgAlatrp	312

1474 ATGTTCCAAAACTTCTTACCCGTAAGATTCGAGAAACATCTTCTAGCTATA 1533
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db LeuPheGluHisPheLeuHisProTyrProSerAspValAspLysHisIleLeuAlaArg 332

1534 CGAAGTCGGCTTGACAAGAAGTCAGGTATCAAACTGGTTTATTAATTCGCGGTTAGGCTA 1593
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db GlnThrGlyLeuSerArgGlyGlnValSerAsnTrpPheIleAsnAlaArgValArgLeu 352

1594 TGGAAAGCCGATGATAGAGATGTATCGGAAATCAACAAGAGGAAGCTCAATAAC--- 1650
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db TrpLysProMetValGluGluMetTyrLeuGluValLysAspProGluAsnAlaIle 372

1651 -----ATCATTCATCAACCAACCGGA----- 1671
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db AlaSerSerGluGlyAlaThrAspGlnAspAsnAspIleAsnProAsnAsnValGlnTyr 392

1672 --CCAACCTCTCGAATGCCAAATCTGTATGATGAGCAACCAATGCATAAATAAGAC 1728
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db ProProProLeuSerSerArgSerGluAspGlnLysProSerLeuValArgIleAsp 412

1729 AACAAATGTGTTTACCACACTTTGTGATAAT 1758
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db SerGluCysAlaSerSerIleIleAsnAsn 422

RESULT 11
US-10-425-114-43080
; Sequence 43080, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jiongong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 43080
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURES:
; OTHER INFORMATION: Clone ID: 700684286_FUL.pep
US-10-425-114-43080

Alignment Scores:
Pred. No.: 2,99e-34 Length: 481
Score: 440.50 Matches: 97
Percent Similarity: 63.43% Conservative: 40
Best Local Similarity: 44.91% Mismatches: 60
Query Match: 13.41% Indels: 19
DB: 12 Gaps: 5

US-09-423-575-1 (1-1886) x US-10-425-114-43080 (1-481)

QY	1057	GGTGATCTTAATAAATCCGAGGCGGGTTTCGA-----TCTCATTTTCAA	1101
D	6	GlyGluAsnAsnAsnAsnAsnAspGlyGlyLysGlnGlyValGluLeuSerThrAlaGln	25
QY	1102	AGGAGAGCATTAAGAAGCAAGAAACCCATCTCTTGGATCTTCTTCAATTCGGTGGATGAT	1161
D	26	ArgGlnGluLeuGlnMetLysLysSerLysLeuValThrMetLeuAspGluValGluGln	45
QY	1162	CGATATAGTCATTCGCTAGATGAGATTCATACGGTTATATCAGGTTTCCATGCTGCAACC	1221
D	46	ArgTyrArgGlnTyrHisGlnMetGlnValValIleThrSerPheGluGlnAlaIle	65
QY	1222	GAGTTAGAT---CCACAGTTACACACCGGTTTCCCTCCAAACCGTTTCTCTTCTTATAC	1278

Db 398 AlaTrrPheGluHisPheLeuHisProTyrProLysAspSerGluLysLeuMetLeu 417
 QY 1528 GCTATACGAAGTGGTGTGCAAGAAGTCAAGTATCAAACTGTTTATAAATCGCGGGT 1587
 Db 418 AlaArgGlnThrGlyLeuThrArgSerGlnIleSerAsnTrpPheIleAsnAlaArgVal 437
 QY 1588 AGCTATGGAAGCGGATGATAGAGAGATGTCGGAAATGAACAAGAGGAGCTCAAT 1647
 Db 438 ArgLeuTrpLysProMetIleGluAspMetTyr-----LysGluIleGly 453
 QY 1648 AACAGTCACATCAACCCACGACCAACTCTCGAATGCCAAATCT 1695
 Db 454 GluAlaAspLeuAspSerAsnSerSerSerAspAsnValProArgSer 469
 RESULT 13
 US-10-425-114-48619
 ; Sequence 48619, Application US/10425114
 ; Publication No. US20040034898A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E.
 ; APPLICANT: Tabaska, Jack E.
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(5313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 48619
 ; LENGTH: 335
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: LIB3592-001-B7_FLI.pep
 US-10-425-114-48619
 Alignment Scores:
 Pred. No.: 5,71e-34 Length: 335
 Score: 437.00 Matches: 105
 Percent Similarity: 54.08% Conservative: 54
 Best Local Similarity: 35.71% Mismatches: 109
 Query Match: 13.31% Indels: 26
 DB: 12 Gaps: 5
 US-09-423-575-1 (1-1886) x US-10-425-114-48619 (1-335)
 QY 802 TCGGAGATAGTCTTTGTCAGCTACAGATAGCTCAGAGGAGGCTTCTTCAGCAGC 861
 Db 9 AlaGlnAlaGlyMetAlaAlaAlaPheLeuSerProAspGlnAlaSerAlaAlaAla 28
 QY 862 AAAGACATTTCTAATAACGTTTCTACTCAAGGTTCTCTCAACTATATATGGCTCAAAA 921
 Db 29 SerArgSerAlaGlnSerIleTyrValGln-----AsnSerArg 41
 QY 922 TACCTTCACTCTGTTCAGAATACTACTCTCATTTTCGCC-----GCA 963
 Db 42 TyrLeuLysAlaAlaArgGluLeuLeuAspGluValValAsnValGlnAspAlaIleLys 61
 QY 964 TACTCGCTCGATTATTCATCTCGAGGACCGAGCAGGCTGCTAGTTCAGCCCTTTACT 1023
 Db 62 ArgLysGlyAspLysSerGlnGlnGlyLysAspSerGlyGlyGlyGlyGlyGlu 81
 QY 1024 TCAGCTTTTGAGATATACCTAGTCTTCTTGATGCTGAT---TCTAATACTCGGAGCG 1080
 Db 82 GlyLysAspAlaGluThrSerAspGluLysAlaGlyGluHisGluGlyAsnSerAla 101
 QY 1081 GGTTCGGATCTACATTTCAAGGAGGAGCATAGAACCAAGAAACCCATCTCTTGGAT 1140
 Db 102 ProGluLeuSerProSerGluArgGlnAspLeuGlnAsnLysValSerAlaLeuMetAla 121

QY 1141 CTCTTCAAATGGTGGATGATCATATAGTCATTGGTAGATGAGATTATACGGTTATA 1200
 Db 122 LeuLeuAspGlnValAspArgLysTyrArgHisGlnMetGlnMetValMet 141
 QY 1201 TCAGCGTTCCATCTGCTCAACCGAGTTAGAT---CCACAGTTACACACCGGTTTCCCTC 1257
 Db 142 SerSerPheAspAlaValAlaGlyAlaGlyAlaAlaArgProTyrThrAlaLeuAlaLeu 161
 QY 1258 CAAACCGTTTCTCTTCTATACAAACCTGAGAGAGAAATCTGCAAGAGATAATCTCT 1317
 Db 162 GlnThrIleSerArgHisPheArgSerLeuArgAspAlaValGlyAlaGlnValGlnSer 181
 QY 1318 ATGGGATCTGTATGGAGAGAGGCAAGCAAGCAAGCTCAAGAAACCTCTATGTTCCACCAG 1377
 Db 182 LeuArgSerLeuGlyLysAspGlySerAlaGlnGlyGlyGlyLeuProArgLeu 201
 QY 1378 CATTGCCCTTCTCAGCAGCTGAAACGAAAGAAC----- 1410
 Db 202 ArgTyrIleAspGlnGlnLeuArgGlnGlnArgAlaMetGlnGlnPheGlyMetMetGln 221
 QY 1411 -----CATCAGATTGGAGACCTCAACGAGTTTCCCTGAGAAATCTCTTCGGTTCTA 1464
 Db 222 GlnProGlnHisAlaTrpArgProGlnArgGlyLeuProGluSerAlaValSerValLeu 241
 QY 1465 CGGAATTTGGATGTTCCAAACTTCTTCCCTTACCCCTTACCCGAAAGATTCCGGAGAAACATCT 1524
 Db 242 ArgAlaTrpLeuPheLysPheLeuHisProTyrProLysAspSerGluLysLeuMet 261
 QY 1525 CTAGCTATACGAAGTGGCTTGACAAAGCTCAGGATCAAGTATCAAACTGGTTTATAAATCGCGG 1584
 Db 262 LeuAlaArgGlnThrGlyLeuSerArgGlyGlnValSerAsnTrpPheIleAsnAlaArg 281
 QY 1585 GTTAGGCTATGGAAGCGGATGATAGAGAGATGATGCGGA 1626
 Db 282 ValArgLeuTrpLysProMetIleGluGluMetTyrArgGlu 295
 RESULT 14
 US-10-437-963-143490
 ; Sequence 143490, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(5322)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 143490
 ; LENGTH: 647
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MET4530_44394C.1.pep
 US-10-437-963-143490
 Alignment Scores:
 Pred. No.: 2,65e-33 Length: 647
 Score: 431.50 Matches: 647
 Percent Similarity: 44.82% Conservative: 89
 Best Local Similarity: 28.06% Mismatches: 218
 Query Match: 13.14% Indels: 75
 DB: 16 Gaps: 17
 US-09-423-575-1 (1-1886) x US-10-437-963-143490 (1-647)

	QY	262	GTTTCTATCAGAAGAAAGGATAAAGTTTCATAGAAAACCCCAATGAACAACAAACAACAC	321
Dy	Db	9	GlySerGluArgAspSerGlnThrMetTyfSerArgAspPro-----GlySerAlaSer	26
	QY	322	AACAACACATTGTAGTTCTCTGGATAATATGCATGACTAACCAAATCCTCTTCTCATGGAT	381
Dy	Db	27	TyrProMetSerSerAlalaLeuGlyAsnLeuLeuTyfLeuAsnAsh-----	41
	QY	382	TTTTATACCCTTCAAGAGAAGATTCCAATCTCAACAAATGCTTCCATCGGAATACCATC	441
Dy	Db	42	-----ProSerGlyProTyfThrGluPheSerGlyIleLeu-----GlnPro	56
	QY	442	AGATCAGATCCTCTACAAATGGTGCTTTGATATTTTCAATTTCTATGTGACTAACAAAA	501
Dy	Db	57	GlnGlnAsnCysMetGluMetProGlyProGlyHisAlaSerAlaMetSerGlnAspPro	76
	QY	502	TACTTATCATCTTCTCCACGGTCTATCGATGTTTCAAGATAAACCGCAATGTGAGTTCATG	561
Dy	Db	77	SerSerArgGluSerAspMetLeuSerSerHisGlnGlyGlnArgSerPheSerHisVal	96
	QY	562	GCTCCTCCTCCTCATCCTCTCCACTTCATCTCTTTGGATCATTTTAAGACAC-----	612
Dy	Db	97	LysAspMetLysasngLuMetLeuMetHisMetMetAspGlyalagInglySerGlySer	116
	QY	613	-----TATCATGATTCCTCAAAACACATCGGGGTTTTGAACCAATATAGTAGTTT	663
Dy	Db	117	GluLeuIleHisaspAlaHisTrGly-----SerGlnLeuGluPhe	131
	QY	664	CAGCATTTTCAGGTGTAGTTGTGCCAAGHGAACCAATGATGTCTACATTCGGTGAAGAA	723
Dy	Db	132	GlyValLeuAsnAshHisasnSerserValproSerMetGlnserGlnGly-----	149
	QY	724	GATTTCOCGTTTCTAAATTCGAATAAAGAAACAATGAGCTTTCATTGAGTCTTGATCA	783
Dy	Db	150	-----LeuSerLeuSerLeuAsnThr	156
	QY	784	GATGTTTTCGATGAATGCTCGGAGATAAGTCTTTGTGCAGCTCACAGATTAGCC-----	837
Dy	Db	157	GlnileMetAlaproSerLeuProTyfTripsSerIleLysProAspMetLeuThrProGln	176
	QY	838	TCAGACCAAGCTTCTGCACGACCAAGACATTTCTAATAACCGTGTACTCAAGGTTTC	897
Dy	Db	177	SerTyfHisaspAsnLeuArgGlygluAspMetArgMetLysasnLeuGlnSerGluAla	196
	QY	898	TCTCAACTTATATTTGGCTCAAAATACCTTCACTCTGTTCACAAGAAATACTATCTATTTC	957
Dy	Db	197	SerArgAlaileArgAsnSersArGtyrLeuLysAlaalaglIngluLeuLeuAspGluVal	216
	QY	958	CCGCATACTCGCTCGATTATTCATCTCAGGA-----ACCGAGTCCAGGACT	1005
Dy	Db	217	ValSerValTrpLysSerIIeLysGlnLysAlaalgnLysgluLysValGluSerGlyLys	236
	QY	1006	GCYAGTTTCAGCCTTACTTCA-----CGTTTTGAGAATAATACTAGTATTTCTTGAT	1056
Dy	Db	237	AlaaspGlyLysGluThrAspGlyGlyProLysSerGlugLyValSer-----	252
	QY	1057	GGTGATTCTTAATAACTCGAGCGCGGTTTCGGA-----TCTACATTTCAA	1101
Dy	Db	253	-----SerAsnProGlnGluSerGlyAlaasnAlaalaproGluLeuSerThrAlaGlu	270
	QY	1102	AGGAGACATTTAGAACCAAGAAACCCATCTCTTGGATCTTCTTCAATGGTGGATGAT	1161
Dy	Db	271	LysGlnGluLeuGlnAsnLysMetAlalyBLeuMetAlametLeuaspGluValaspArg	290
	QY	1162	CGATATAGTCAATTCGTAGATGAGATTACACGGTTTATACAGCGTTTCCAT---GOTGCA	1218
Dy	Db	291	LysTyfLysHisTyfTyfHisGlnMetGlnThrValSerSerPheAspValValAla	310
	QY	1219	ACCGAGTTAGATCCACAGTTACACACCGGTTTCGCTCCAAAACCGTTTCTCTCTATAC	1278
Dy	Db	311	GlyProGlySerAlaLysProTyfThralaValAlaleuGlnThrIIIeSerArgHisPhe	330
	QY	1279	AAGAACTCTGAGAGAGAGAAATCTGCAAGAAGAAATATCTCTATGGGATCTGATTG---GAG	1335

Db	331	ArgCysLeuYsAspAlaIleAsnAspGlnIleAsnValIleArgLysLysLeuGlyGlu	350
Qy	1336	AGAGGCAAGACAAAGACTCAAGAAACCTATGTTCACACGAGATTGCCCTTCTTCCAGCAG	1395
Db	351	GluGluAsnSerSerGlyLysGluGlyLysLeuThrArgLeuArgTyrIleAspGlnGln	370
Qy	1396	CTGAACCGAAGAAC-----CATCAGATTTCGAGACCT	1428
Db	371	LeuArgGlnGlnArgAlaPheGlnGlnTyrGlyMetIleProGlnAsnAlaTrpArgPro	390
Qy	1429	CAACGAGGTTGCCTCAGAAATCTGTTTCGGTCTTACGGAATTCGATGTTCCAAAACCTTC	1488
Db	391	GlnArgGlyLeuProGluAsnSerValThrIleLeuArgAlaTrpLeuPheGluHisPhe	410
Qy	1489	CTTCACCCCTACCCGAAAGATTCCGAGAACATCTCTTACGTATACGAAGTGCCTTGACA	1548
Db	411	LeuHisProTyrProLysAspSerGluLysLeuMetLeuAlaArgGlnThrGlyLeuThr	430
Qy	1549	AGAAGTCAGGTATCAAACTGGTTTATAAATCCGCGGTTAGGCTATCGAAGCCGATGATA	1608
Db	431	ArgSerGlnIleSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIle	450
Qy	1609	GAGAGATGTATCGGAAATGAACACAGAGAACTCAATAACAGTCACATTCACCCCAAC	1668
Db	451	GluAspMetTyr-----LysGluGluIleGlyAspLeuGluGlnAspSerAsn	466
Qy	1669	GGACCAACTCTCGAATGCCAAATCTGTT---ATGATGACCAAGCAATGCATAAATAA	1725
Db	467	SerSerSerAspAsnAlaProArgSerLysAspLysMetAlaSerSerGluAspLysGlu	486
Qy	1726	GAC---AACAACTGTGTTTACCACCTTTGTGAT	1755
Db	487	AspLeuLysAsnSerArgAlaAlaGlyCysGlu	497

RESULT 15

US-10-424-599-193114

; Sequence 193114, Application US/10424599

; Publication No. US200400031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with Soybean

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 193114

; LENGTH: 133

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

US-10-424-599-193114

Alignment Scores:			
Pred. No.:	2,68e-33	Length:	133
Score:	428.50	Matches:	82
Percent Similarity:	74.19%	Conservative:	10
Best Local Similarity:	66.13%	Mismatches:	17
Query Match:	13.05%	Indels:	15
DB:	12	Gaps:	1

US-09-423-575-1 (1-1886) x US-10-424-599-193114 (1-133)

Qy 1342 AAACACAAAGACTCAAGAAACCTCTATGTTCCACCAGCATTGCCCTTCTTCCAGCAGTGA

1401

Db 10 LysAspArgThrPheGluThrSerPheIleArgGluGlnTrpAlaLeuGlnGlnLeuArg

29

Qy 1402 CGAAAGAACCATCAGATTGGAGACCTCAACGAGGTTGCCCTGAGAAATCTGTTTCGGT

1461

30	ArgAsnGluGlnSerTrpArgProGlnArgGlyLeuProGluLysSerValSerVal	49
1462	CTACCGAATTGGATGTTCCAAAACCTTCACCCCTACCCGAAGATTCGGAGAACAT	1521
50	LeuArgAlaTrpMetPheGlnAsnPheLeuHisProTyrProLysAspAsnGluLysHis	69
1522	CTTCTAGCTATACGAAGTGGCTTGACAAGAAGTCAGGTATCAAACTGGTTTAAATGGC	1581
70	LeuLeuAlaIleGlnSerGlyLeuThrArgSerGlnValSerAsnTrpPheIleAsnAla	89
1582	CGGTTTAGCTATAGGAACCGCATGATAGAAGAGATGTATCCGAATATGAACAAGG---	1638
90	ArgValArgLeuTrpLysProLeuIleGluGluMetTyrSerGluValAsnLysLysAsn	109
1639	-----AAGCTCAATAACACAGTCAC	1656
110	ArgProIleGluGlySerGlyCysAspSerArgAsnTyrAlaSerIleGlyAsnGlnArg	129
1657	ATTCAACCCCAAC	1668
130	IleGlnMetAsn	133

Search completed: September 2, 2004, 12:39:23
Job time : 356.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2004, 00:11:20 ; Search time 754 Seconds
(without alignments)
10626.129 Million cell updates/sec

Title: US-09-423-575-1
Perfect score: 1886
Sequence: 1 attagttataaatgttc.....ttaaaaaaaaaaaaaa 1886

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 337863 seqs, 212409041 residues 674726

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04: *
1: geneseqn1980s: *
2: geneseqn1990s: *
3: geneseqn2000s: *
4: geneseqn2001as: *
5: geneseqn2001bs: *
6: geneseqn2002s: *
7: geneseqn2003as: *
8: geneseqn2003bs: *
9: geneseqn2003cs: *
10: geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1882.8	99.8	1886	2	AAX01422
2	1476	78.3	9359	6	ABK11039
3	1476	78.3	9359	6	ABK10687
4	128.4	6.8	323	3	AAC56287
5	127.4	6.8	343	3	AAC56856
6	126.2	6.7	2025	3	AAC39138
7	123.6	6.6	316	3	AAC57132
8	123.6	6.6	462	3	AAC56228
9	123.6	6.6	2043	6	ABZ13532
10	123.6	6.6	2043	8	ADB23149
11	123.6	6.6	2385	9	ADD30303
12	122.2	6.5	529	3	AAC55946
13	121.8	6.5	1539	6	ABZ14349
14	121.2	6.4	833	3	AAC34112
15	119.8	6.4	1449	3	AAC43406
16	119.8	6.4	2031	3	AAC36460
17	114.4	6.1	1884	6	ABZ13100
18	114.4	6.1	1983	4	AAD06494
19	114.4	6.1	1983	9	ADC46618
20	114.4	6.1	1983	9	ADD30786
21	113.4	6.0	2545	4	AAD06479
22	113.4	6.0	2545	9	ADE37296
23	112.2	5.9	2450	7	AAL55527

24	110.6	5.9	2442	7	AAL55526	AAL55526 qSH-1 gen
25	110	5.8	1836	6	ABZ12846	Abz12846 Arabidops
26	110	5.8	2405	4	AAD06449	Aad06449 Arabidops
27	110	5.8	2405	9	ADE37116	Ade37116 Plant yie
28	107	5.7	404	7	ABX62621	Abx62621 Arabidops
29	107	5.7	1575	4	AAD06493	Aad06493 Arabidops
30	107	5.7	1824	3	AAC36745	Aac36745 Arabidops
31	103.2	5.5	1905	3	AAC39451	Aac39451 Arabidops
32	100.6	5.3	2131	6	ABL56175	AbL56175 Hordeum v
33	100.6	5.3	2141	6	ABL56177	AbL56177 Hordeum v
34	96.8	5.1	381	3	AAC55971	Aac55971 Eucalyptu
35	96.8	5.1	381	3	AAC56750	Aac56750 Eucalyptu
36	94.4	5.0	380	3	AAC56524	Aac56524 Eucalyptu
37	94.4	5.0	380	3	AAC55980	Aac55980 Eucalyptu
38	92.4	4.9	626	5	AAH87740	Aah87740 Peppermin
39	89.8	4.8	706	6	ABK82114	Abk82114 DNA encod
40	84.8	4.5	260	3	AAC56220	Aac56220 Pinus rad
41	84.8	4.5	260	3	AAC57111	Aac57111 Pinus rad
42	79.8	4.2	366	3	AAC56519	Aac56519 Eucalyptu
43	79.2	4.2	474	3	AAC56677	Aac56677 Eucalyptu
44	79.2	4.2	474	3	AAC55926	Aac55926 Eucalyptu
45	79	4.2	157	3	AAC56263	Aac56263 Pinus rad

ALIGNMENTS

RESULT 1
AAX01422
ID AAX01422 standard; cDNA; 1886 BP.
XX
AC AAX01422;
DT 26-APR-1999 (first entry)
XX
DE A. thaliana ATH1 homeobox gene.
XX
KW ATH1; homeobox gene; plant gene construct; flowering modification;
KW promoter; shade avoidance response; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO9851800-A1.
XX
PD 19-NOV-1998.
XX
PF 14-MAY-1998; 98WO-IB000821.
XX
PR 14-MAY-1997; 97GB-00009789.
PR 30-DEC-1997; 97GB-00027458.
XX
XX (VHAV-) VAN DER HAVE BV D J.
XX
XX Smeekens S, Weisbeek P, Proveniers M;
XX
XX WPI; 1999-070121/06.
XX
XX Plant gene constructs containing ATH1 - used to modify flowering or the
XX shade avoidance response in plants.
XX
XX Claim 5; Fig 1; 53pp; English.
XX
XX This sequence represents the Arabidopsis thaliana ATH1 homeobox gene. The
XX invention relates to a plant gene construct comprising a complete or
XX partial DNA sequence coding for an ATH1 gene product under the control of
XX a promoter functional in plants. The construct can be used to modify
XX (either promote or retard) flowering in plants. The construct can also be
XX used in a method of producing a transgenic plant that lacks the shade
XX avoidance response of a wild type plant

SQ Sequence 1886 BP; 599 A; 373 C; 344 G; 570 T; 0 U; 0 Other;
Query Match 99.8%; Score 1882.8; DB 2; Length 1886;

		Best Local Similarity 99.9%; Pred. No. 0;				Matches 1884; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	1	ATTAGTATATAAATGTTCTATTTTGTGATCTAGTCTCTGAATCTTTTGTAGTGAGGAG	60						
DB	1	ATTAGTATATAAATGTTCTATTTTGTGATCTAGTCTCTGAATCTTTTGTAGTGAGGAG	60						
QY	61	ATGATGAAGATATGAATTTCTTCAATGAAATTTTGTAAAGAAAGAAACATAGAGAGCT	120						
DB	61	ATGATGAAGATATGAATTTCTTCATGAATTTTGTAAAGAAAGAAACATAGAGAGCT	120						
QY	121	GCGAATGAAGTACACTGTTCTTTTCAGGAGAAAGAAAGATTAATTAAGCATTTCTTT	180						
DB	121	GCGAATGAAGTACACTGTTCTTTTCAGGAGAAAGAAAGATTAATTAAGCATTTCTTT	180						
QY	181	CTTCAGTTTAAACACACATTTTGGAAATTTTGTATGATTAATTTCTTTTGGAAAGTTGT	240						
DB	181	CTTCAGTTTAAACACACATTTTGGAAATTTTGTATGATTAATTTCTTTTGGAAAGTTGT	240						
QY	241	GTGTCTGAATCTTCCCAAGTTCTATCAAGAGAAAGATTAATTTCTATAGAAAC	300						
DB	241	GTGTCTGAATCTTCCCAAGTTCTATCAAGAGAAAGATTAATTTCTATAGAAAC	300						
QY	301	CAATGGACAAACAAACAAACAAACAACTTTTGTCTCTGGATTAATGTCACTAAC	360						
DB	301	CAATGGACAAACAAACAAACAAACAACTTTTGTCTCTGGATTAATGTCACTAAC	360						
QY	361	CAAAATCTCTCTCATGATTTTATACCTTCAAGAGAAAGATTTCAACTTTCTCAACA	420						
DB	361	CAAAATCTCTCTCATGATTTTATACCTTCAAGAGAAAGATTTCAACTTTCTCAACA	420						
QY	421	ATGCTTCATGAATACATCATGATCATGATCTTCAAAATGGGGTCTTTGATATTTTC	480						
DB	421	ATGCTTCATGAATACATCATGATCATGATCTTCAAAATGGGGTCTTTGATATTTTC	480						
QY	481	AATTCATCTCTCACTAAACAAATTTATCATCTTCTCCACGGTCTATCGATGTTCAAGAT	540						
DB	481	AATTCATCTCTCACTAAACAAATTTATCATCTTCTCCACGGTCTATCGATGTTCAAGAT	540						
QY	541	AACCGAATGTTGAGTTCATGGCTCTCTCTCTCATCTCTCTCACTTTCACTCTTTGGAT	600						
DB	541	AACCGAATGTTGAGTTCATGGCTCTCTCTCTCATCTCTCTCACTTTCACTCTTTGGAT	600						
QY	601	CATTTAAGACATATGATGATTTCTTCAACACATGTTGGGTTTGAAGCAATAGTGAG	660						
DB	601	CATTTAAGACATATGATGATTTCTTCAACACATGTTGGGTTTGAAGCAATAGTGAG	660						
QY	661	TTTCAGGCATTTTCAGGTGTAGTTGGTCCAAAGTGAACCAATGATGCTTACATTCGGTGAA	720						
DB	661	TTTCAGGCATTTTCAGGTGTAGTTGGTCCAAAGTGAACCAATGATGCTTACATTCGGTGAA	720						
QY	721	GAAGATTTCCGTTCTTAATTTGATTAAGAAACATGAGCTTCAATGAGTCTTGCA	780						
DB	721	GAAGATTTCCGTTCTTAATTTGATTAAGAAACATGAGCTTCAATGAGTCTTGCA	780						
QY	781	TCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTTGTGAGCAATAGAGCTCA	840						
DB	781	TCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTTGTGAGCAATAGAGCTCA	840						
QY	841	GAGCAAGCTTTTCAGGAGCAAGACATTTCTAATACGTTGTTACTCAAGGTTTCTCT	900						
DB	841	GAGCAAGCTTTTCAGGAGCAAGACATTTCTAATACGTTGTTACTCAAGGTTTCTCT	900						
QY	901	CAACTTATTTTGGCTCAAAATACCTTCACTCTGTTCAAGAAATACTATCTCATTTGCC	960						
DB	901	CAACTTATTTTGGCTCAAAATACCTTCACTCTGTTCAAGAAATACTATCTCATTTGCC	960						
QY	961	GCATACCTCGCTGATTAATTCATCTCGAGAAACCGAGTCAGAGCTGCTAGTTCAGCTTT	1020						
DB	961	GCATACCTCGCTGATTAATTCATCTCGAGAAACCGAGTCAGAGCTGCTAGTTCAGCTTT	1020						
QY	1021	ACTTCACGTTTTCAGCAATATACCTGATTTCTTGTGATGATTTCTAATACTCGAGGCG	1080						

DB	1021	ACTTCACGTTTTCAGCAATATACCTGATTTCTTGTGATGTTCTAATACTCGAGGCG	1080
QY	1081	GGTTTCGGATCTACATTTTCAAGGAGAGCATTTAGAGCAAGAAACCCATCTCTTGGAT	1140
DB	1081	GGTTTCGGATCTACATTTTCAAGGAGAGCATTTAGAGCAAGAAACCCATCTCTTGGAT	1140
QY	1141	CTTCTTCAAAATGCTGATGATCGATAGTCAATGCTAGATGAGATTTCATCGGTTATA	1200
DB	1141	CTTCTTCAAAATGCTGATGATCGATAGTCAATGCTAGATGAGATTTCATCGGTTATA	1200
QY	1201	TCAGCGTTCCATCTCTGCAACCGAGTTAGATCCACATTCACACCCCGTTTCCCTCCAA	1260
DB	1201	TCAGCGTTCCATCTCTGCAACCGAGTTAGATCCACATTCACACCCCGTTTCCCTCCAA	1260
QY	1261	ACCGTTTCCCTTCTTATACAAAGAACCTGAGAGAGAAATCTGCAAGAAAGATAATCTCTATG	1320
DB	1261	ACCGTTTCCCTTCTTATACAAAGAACCTGAGAGAGAAATCTGCAATATATATCTCTATG	1320
QY	1321	GGATCTGTATTTGAGAGAGGCAAGACAAGACTCAAGAAACCTCTATGTTCCACAGCAT	1380
DB	1321	GGATCTGTATTTGAGAGAGGCAAGACAAGACTCAAGAAACCTCTATGTTCCACAGCAT	1380
QY	1381	TGCTTTCTTTCAGCAGCTGAAACGAAAGAACCTCAGATTTGGAGACCTCAACGAGGTTG	1440
DB	1381	TGCTTTCTTTCAGCAGCTGAAACGAAAGAACCTCAGATTTGGAGACCTCAACGAGGTTG	1440
QY	1441	CCTGAGAAATCTGTTTCGGTTCTTACGGAATTTGGATTTCCAAAACCTTCCACCTTAC	1500
DB	1441	CCTGAGAAATCTGTTTCGGTTCTTACGGAATTTGGATTTCCAAAACCTTCCACCTTAC	1500
QY	1501	CCGAAAGATTCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAGAAAGTCAGTA	1560
DB	1501	CCGAAAGATTCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAGAAAGTCAGTA	1560
QY	1561	TCAAACTGGTTTATAAATCGCGGGTTAGGCTATGGAAGCCGATGATAGAAGATGTAT	1620
DB	1561	TCAAACTGGTTTATAAATCGCGGGTTAGGCTATGGAAGCCGATGATAGAAGATGTAT	1620
QY	1621	GCGAAATGAACAAAGAGGAGCTCAATAACAGTCACATTCACCCAAACCGACCACTCTT	1680
DB	1621	GCGAAATGAACAAAGAGGAGCTCAATAACAGTCACATTCACCCAAACCGACCACTCTT	1680
QY	1681	CGAATCCCAAAATCTCTTATGATGAGCCAAAGCAATGCAATAAGACAAACATTTGTTT	1740
DB	1681	CGAATCCCAAAATCTCTTATGATGAGCCAAAGCAATGCAATAAGACAAACATTTGTTT	1740
QY	1741	TACCAACTTTGTGATAAATTAGCAATTTGCTACTCTATGATTTGCCCAAAACCTTAAACCATG	1800
DB	1741	TACCAACTTTGTGATAAATTAGCAATTTGCTACTCTATGATTTGCCCAAAACCTTAAACCATG	1800
QY	1801	TAGCACTATCATTAGTATGTTTATATGTTATATACACTCTCTTATCTTTGACTATTTTC	1860
DB	1801	TAGCACTATCATTAGTATGTTTATATGTTATATATACACTCTCTTATCTTTGACTATTTTC	1860
QY	1861	ATTTTTAAAAAATAAAAAAAAAAAAAA	1886
DB	1861	ATTTTTAAAAAATAAAAAAAAAAAAAA	1886

RESULT 2
 ABK11039
 ID ABK11039 standard; DNA; 9359 BP.
 XX
 AC ABK11039;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE pVDH636 vector containing Arabidopsis Ath1 gene.
 XX
 KW pVDH636 vector; circular; cyclic; grass; plant; herbicide resistance;
 KW baseball; cricket; football; golf; rugby; soccer; tennis; lawn; park;
 KW athletic field; animal feedstuff; grass flowering inhibitor; Ath1; gene;
 KW db.

QY 1619 ATGCGGAATGACAGAGAGAGCTCAATAACAGTCACATTCACGCCAAGCAGCAACTC 1678
 Db 6245 ATGCGGAATGACAGAGAGAGCTCAATAACAGTCACATTCACGCCAAGCAGCAACTC 6304
 QY 1679 TTCGAATGCCAAATCTGTTATGATGAGCAAGCAATGATATAAAGACAACTTTG 1738
 Db 6305 TTCGAATGCCAAATCTGTTATGATGAGCAAGCAATGATATAAAGACAACTTTG 6364
 QY 1739 TTTACCAACTTTGTGATAATTAGGAATTGCTACTC 1774
 Db 6365 TTTACCAACTTTGTGATAATTAGGAATTGCTACTC 6400

RESULT 3

ABK10687
 ID ABK10687 standard; DNA; 9359 BP.
 AC ABK10687;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE Transformation vector plasmid pVDH636 DNA.
 XX
 KW Grass; plant; transgenic; flowering inhibition; inflorescence; gene; ds;
 KW tiller production; delayed heading; gibberellic acid; phytohormone; golf;
 KW genetically modified grass; athletic field; sport; baseball; cricket;
 KW football; rugby; soccer; tennis; lawn; landscaping; cattle; horse; sheep;
 KW goat; animal feedstuff; Arabidopsis thaliana; Ath1; homeobox gene;
 KW circular; cyclic; pVDH636.
 XX
 OS Synthetic.

XX Key Location/Qualifiers
 FH 4930..6351
 FT CDS
 FT /*tag= a
 FT /product= "Arabidopsis thaliana Ath1"

XX WO200214486-A2.
 XX
 XX 21-FEB-2002.
 XX
 XX 16-AUG-2001; 2001WO-EP009570.
 XX
 XX 18-AUG-2000; 2000US-0226422P.
 PR 27-NOV-2000; 2000US-0253274P.
 PR 27-NOV-2000; 2000US-0253327P.
 PR 22-JUN-2001; 2001US-0300220P.
 XX
 XX (ADVA-) ADVANTA SEEDS BV.
 XX
 XX Van Der Valk P, Van Dun CMP, Smeekens SCM, Proveniers MCG;
 PI P-PSDB; AAU76885.
 XX
 DR WPI; 2002-257603/30.
 DR P-PSDB; AAU76885.
 XX
 PT New genetically modified grass useful for growing and/or propagating
 PT grass in athletic fields (for sports such as baseball, cricket, football,
 PT golf, rugby, soccer and tennis), or as animal feedstuff for cattle, goat,
 PT horse and sheep.
 XX
 XX Example 1; Fig 6; 45pp; English.
 PS
 XX The invention relates to a grass plant which has been genetically
 CC modified to substantially inhibit generative propagation. The genetic
 CC modification may result in a heritable change in one or more plant
 CC characteristics such as inhibition of flowering (or substantial delay
 CC that amounts to inhibition), absence of inflorescence, increased
 CC production of tillers, delayed heading and inhibition of the
 CC developmental switch from vegetative to generative growth. A method of
 CC making a grass involves transforming the grass with a nucleic acid which
 CC interferes with metabolism of gibberellic acid. A grass can be treated by
 CC applying a phytohormone to at least partially relieve or reverse a change
 CC in plant characteristic resulting from genetic modification. The

CC genetically modified grass is useful for growing and/or propagating grass
 CC in athletic fields (for sports such as baseball, cricket, football, golf,
 CC rugby, soccer and tennis), lawns, parks and other types of landscaping.
 CC The grass is also useful as an animal feedstuff for cattle, goats, horses
 CC and sheep, due to its increased vegetative growth, improved digestibility
 CC and/or nutritional value as animal feedstuff. This sequence represents
 CC plasmid pVDH636 DNA. This plasmid was used as a transformation vector for
 CC production of transgenic grasses expressing the Arabidopsis thaliana
 CC homeobox gene, Ath1
 XX
 SQ Sequence 9359 BP; 2377 A; 2222 C; 2133 G; 2624 T; 0 U; 3 Other;

Query Match 78.3%; Score 1476; DB 6; Length 9359;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 299 ACCAATGGACAACAACAACAACACTTTTAGTTCTCTGGATAATGTCAGCTA 358
 Db 4925 ACCAATGGACAACAACAACAACACTTTTAGTTCTCTGGATAATGTCAGCTA 4984
 QY 359 ACCAATGCTCTTCTCATGGATTTTATACCTTCAAGAGAAGATTCAACTTCTCAA 418
 Db 4985 ACCAATGCTCTTCTCATGGATTTTATACCTTCAAGAGAAGATTCAACTTCTCAA 5044
 QY 419 CAATGCTTCCATGGAATACCATCAGATCAGATCCTCTACAAATGGTGGCTTTGATATT 478
 Db 5045 CAATGCTTCCATGGAATACCATCAGATCAGATCCTCTACAAATGGTGGCTTTGATATT 5104
 QY 479 TCAATTCTATGCTGACTAACAAATACTTATCATCTTCTCCAGGCTATCGATGTCAG 538
 Db 5105 TCAATTCTATGCTGACTAACAAATACTTATCATCTTCTCCAGGCTATCGATGTCAG 5164
 QY 539 ATAAACCGCAATGTTAGTTTCATGGCTCCCTCCCTCATCTCTCCACATTCATCTTTGG 598
 Db 5165 ATAAACCGCAATGTTAGTTTCATGGCTCCCTCCCTCATCTCTCCACATTCATCTTTGG 5224
 QY 599 ATCAATTAAGACACTATGATGTTCTCAAACAACATGTGGGGTTTGAAGCAAAATAGTG 658
 Db 5225 ATCAATTAAGACACTATGATGTTCTCAAACAACATGTGGGGTTTGAAGCAAAATAGTG 5284
 QY 659 AGTTTCAGGCATTTTCAGGTGATGTTGTCGAAGTGAACCAATGATGTACATTCGGTG 718
 Db 5285 AGTTTCAGGCATTTTCAGGTGATGTTGTCGAAGTGAACCAATGATGTACATTCGGTG 5344
 QY 719 AAGAAGATTTCGGTTTCTAAATTCGAATTAAGAAGAACAAATCAGCTTTTCATGAGTCTTG 778
 Db 5345 AAGAAGATTTCGGTTTCTAAATTCGAATTAAGAAGAACAAATCAGCTTTTCATGAGTCTTG 5404
 QY 779 CATCAGATGTTTCTGATGATGCTCGGAGATAAGTCTTTGTGCAGCTACAAGATTAGCCT 838
 Db 5405 CATCAGATGTTTCTGATGATGCTCGGAGATAAGTCTTTGTGCAGCTACAAGATTAGCCT 5464
 QY 839 CAGAGCAAGCTTCTTCAGCAGCAAGACATTTCTAATAACGTTGTACTCAAGTCTTCT 898
 Db 5465 CAGAGCAAGCTTCTTCAGCAGCAAGACATTTCTAATAACGTTGTACTCAAGTCTTCT 5524
 QY 899 CTCACCTTATATTTGCTCAAAATACCTTCATCTGTTTCAAGAAATACATCTCTATTG 958
 Db 5525 CTCACCTTATATTTGCTCAAAATACCTTCATCTGTTTCAAGAAATACATCTCTATTG 5584
 QY 959 CCGCATACTCGTTCGATTATTTCATCTCGAGGAACCGAGTCAGGAGCTGTAGTTCAGCCT 1018
 Db 5585 CCGCATACTCGTTCGATTATTTCATCTCGAGGAACCGAGTCAGGAGCTGTAGTTCAGCCT 5644
 QY 1019 TTAATTCAGCTTTTGAGATAATACTGAGTTTCTTGTGATGCTGATTTCTTAATACTCGAGG 1078
 Db 5645 TTAATTCAGCTTTTGAGATAATACTGAGTTTCTTGTGATGCTGATTTCTTAATACTCGAGG 5704
 QY 1079 CGGGTTTCCGATCTACATTTCAAAGGAGAGCATTTAGAAGCAAGAAACCCATCTCTTGG 1138
 Db 5705 CGGGTTTCCGATCTACATTTCAAAGGAGAGCATTTAGAAGCAAGAAACCCATCTCTTGG 5764
 QY 1139 ATCTTCTTCAAATGGTGGATGATGATATAGTCATTCGTGTAGATGAGATTTCATACGGTTA 1198

5765 ATCTCTCTCAATGTTGGATGATCGATATAGTCAATTCGTTAGATGAGATTCATACGGTTA 5824
 1199 TATCAGCGTTCATGCTCAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCTCC 1258
 5825 TATCAGCGTTCATGCTCAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCTCC 5884
 1259 AAACCGTTTCTCTTATACAAAGAACCTGAGAGAGAGATCTGCAAGAGAGATAATCTCTA 1318
 5885 AAACCGTTTCTCTTATACAAAGAACCTGAGAGAGAGATCTGCAAGAGAGATAATCTCTA 5944
 1319 TGGGATCTGATTTGAGAGAGAGCAAGACAGACTCAAGAACCTCTATGTTCCACAGC 1378
 5945 TGGGATCTGATTTGAGAGAGAGCAAGACAGACTCAAGAACCTCTATGTTCCACAGC 6004
 1379 ATTGCCTTCTCAGCAGCTGAAACGAAGAACCATCAGATTGGAGACTCAACGAGTT 1438
 6005 ATTGCCTTCTCAGCAGCTGAAACGAAGAACCATCAGATTGGAGACTCAACGAGTT 6064
 1439 TGCCTGAGAAATCTGTTTCGGTTCTACGGAATGGATGTTCCAAAATCTCTTACCCCTT 1498
 6065 TGCCTGAGAAATCTGTTTCGGTTCTACGGAATGGATGTTCCAAAATCTCTTACCCCTT 6124
 1499 ACCCGAAGATTCGGAGAGAAATCTCTAGCTATACGAGTGGCTTGACAGAGTCAAG 1558
 6125 ACCCGAAGATTCGGAGAGAAATCTCTAGCTATACGAGTGGCTTGACAGAGTCAAG 6184
 1559 TATCAAACTGTTTATAATGCGCGGTTAGCTATGGAAGCGGATGATAGAAGATGT 1618
 6185 TATCAAACTGTTTATAATGCGCGGTTAGCTATGGAAGCGGATGATAGAAGATGT 6244
 1619 ATGCGGAATGAAACAGAGAGAGCTCAATAACAGTCAATTAACCCACGACCACTC 1678
 6245 ATGCGGAATGAAACAGAGAGAGCTCAATAACAGTCAATTAACCCACGACCACTC 6304
 1679 TTGGAATGCAAAATCTGTTATGATGAGCGCAAGCAATGCATAATAAAGACAACTTGT 1738
 6305 TTGGAATGCAAAATCTGTTATGATGAGCGCAAGCAATGCATAATAAAGACAACTTGT 6364
 1739 TTACCACTTGTGATATTAAGCAATGCTACTC 1774
 6365 TTACCACTTGTGATATTAAGCAATGCTACTC 6400

RESULT 4
 AAC56287
 ID AAC56287 standard; DNA; 323 BP.
 AC AAC56287;
 XX
 XX 25-JAN-2001 (first entry)
 DE Pinus radiata transcription factor DNA sequence #87.
 XX
 XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.
 XX
 XX Pinus radiata.
 OS
 XX
 XX WO200053724-A2.
 PN
 XX
 XX 14-SEP-2000.
 PD
 XX
 XX 09-MAR-2000; 2000WO-US006112.
 PF
 XX
 XX 11-MAR-1999; 99US-00266513.
 PR
 XX
 XX 18-AUG-1999; 99US-0149485P.
 PP
 XX
 XX (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Wood M, McGrath A, Shenk MA, Glenn M;
 XX WPI; 2000-579369/54.
 XX
 XX New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide.
 XX
 XX Claim 1; Page 153; 747pp; English.
 XX
 XX The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and
 CC MYB
 XX
 XX Sequence 323 BP; 99 A; 67 C; 80 G; 77 T; 0 U; 0 Other;
 SQ
 Query Match 6.8%; Score 128.4; DB 3; Length 323;
 Best Local Similarity 67.7%; Pred. No. 9.5e-22;
 Matches 180; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
 QY 1370 TCACACGAGCTTCCTCTTCCAGCAGCTGAAACGAAGAACCATCAGATTGGAGACTC 1429
 Db 15 TACGACAAACGCGGCTTTCACCATTTAGATGATGAGAGAGACCTTGGGACCGC 74
 QY 1430 AACGAGGTTTCCCTGAGAAATCTGTTTCGGTTCTACGGAATTCGATGTTCCAAAATCTC 1489
 Db 75 AGAGAGGACTTCCTGAAACGCTGTTCTGTTCTGTCATGTTGTTGAGCATTTTC 134
 QY 1490 TTCACCTTACCGAAGATTCGAGAAACATCTTCTAGCTATACGAGTGGCTTGACAA 1549
 Db 135 TGCACCGGTATCCAACTGATGACAGATAAGCATATATTTGGCTAAAGAACTGGCCTTACAA 194
 QY 1550 GAAGTCAGGTATCAAACTGTTTATAAATCGCGGGTTAGGCTATGGAAGCCGATGATAG 1609
 Db 195 GAAGTCAGGTATCAAACTGTTTATAAATCGCGGGTTAGGCTATGGAAGCCGATGATAG 1609
 QY 1610 AAGAGATGATGCGGAATGAAACAAAG 1635
 Db 255 AGGAGATGATGCGGAATGAAACAAAG 280
 RESULT 5
 AAC56856
 ID AAC56856 standard; DNA; 343 BP.
 AC AAC56856;
 XX
 XX 25-JAN-2001 (first entry)
 DE Pinus radiata transcription factor DNA sequence #302.
 XX
 XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.
 XX
 XX Pinus radiata.
 OS
 XX
 XX WO200053724-A2.
 PN
 XX
 XX 14-SEP-2000.
 PD
 XX
 XX 14-SEP-2000.

```

PF 09-MAR-2000; 2000WO-US006112.
XX
XX 11-MAR-1999; 99US-00266513.
PR 18-AUG-1999; 99US-0149485P.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
FA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Shenk MA, Glenn M;
XX WPI; 2000-579369/54.
DR
XX
XX New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide.
XX
XX Claim 1; Page 530; 747pp; English.
PS
XX
XX The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper, LIM domain, AP2
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and
CC MYB
XX
XX
SQ Sequence 343 BP; 106 A; 70 C; 86 G; 79 T; 0 U; 2 Other;

Query Match 6.8%; Score 127.4; DB 3; Length 343;
Best Local Similarity 67.3%; Pred. No. 1.7e-21;
Matches 179; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 1370 TCCACGAGCATTCGCTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACCTC 1429
DB 35 TACGACACAGCGCCGCAATTCACCACTTAGGATTGATGGAGCAGCACCCCTTGGCGACCGC 94

QY 1430 AACGAGGTTTGCTGAGAAATCTGTTTCGGTTCTACGGAATGATGTTCCAAACTTCC 1489
DB 95 AGAGAGGACTTCTCGAAGCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 154

QY 1490 TTCACCTTACCCGAAAGATTCGAGAAACATCTTCTAGCTATACGAGTGGCTTGACAA 1549
DB 155 TCCACCGGTATCCACTGATGAGATAGCATATATTTGGCTTAGCAAACTGGCCTTACAA 214

QY 1550 GAAGTCAGATATCAACTGGTTTATAAATGCGCGGTTAGCTATGGAAGCCGATGATAG 1609
DB 215 GAAGTCAGATATCAACTGGTTTATAAATGCGCGGTTAGCTATGGAAGCCGATGATG 274

QY 1610 AAGAGATGTATCGGAATGACACAG 1635
DB 275 AGGAGATGTATCGGAAGAACTCAAG 300

RESULT 6
AAC39138
ID AAC39138 standard; DNA; 2025 BP.
XX
XX AAC39138;
AC
XX
XX 17-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 23512.
DE
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS

```

```

XX EF1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000BP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 08-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 08-MAY-1999; 99US-0132486P.
PR 08-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136382P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.

```

```
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148344P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 08-OCT-1999; 99US-0157753P.
PR 08-OCT-1999; 99US-0157865P.
PR 08-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159284P.
PR 13-OCT-1999; 99US-0159285P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161320P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 08-OCT-1999; 99US-0157753P.
PR 08-OCT-1999; 99US-0157865P.
PR 08-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159284P.
PR 13-OCT-1999; 99US-0159285P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161320P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
Query Match 6.7%; Score 126.2; DB 3; Length 2025;
Best Local Similarity 68.0%; Pred. No. 6.Se-21; Indels 0; Gaps 0;
Matches 176; Conservative 0; Mismatches 83;
Qy 1368 GTTCCACCAGATTGCTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTTGGAGACC 1427
Db 1196 GTTGAGACGCAAGAGCTTTGCATCACAGCTTGGATGGTTCCGCCCTTGGAGACC 1255
Qy 1428 TCACGAGGTTGCTCGTGAAGAAATCTGTTTCGGTCTACGGAATGGATGTTCCAAACATT 1487
Db 1256 TCAAGAGGTTCTCTCGTGAAGAACTCTGTTTCTCGTCTGGCTCTTCGACATTT 1315
Qy 1488 CCTTCACCTTACCCGGAAGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGAC 1547
Db 1316 CCTTCATCCATATCCGGAAGAAATCTGAGAAATCATGCTTCCGAAACAGACAGACTGTC 1375
Qy 1548 AAGAAGTCAGGTATCAAACTGGTTTATAATGCGCGGTTAGGCTATGGAAGCCGATGAT 1607
Db 1376 GAAGAACGAGTTCTGAATGGTTTATAAAGCGGAGAGTTCTGTTATGGAACCGATGAT 1435
Qy 1608 AGAAGATCTATCGGAA 1626
Db 1436 CGAAGATGTATAAGAA 1454
RESULT 7
AAC57132
ID AAC57132 standard; DNA; 316 BP.
XX
AC AAC57132;
XX
```

```

DT XX 25-JAN-2001 (first entry)
DE XX Pinus radiata transcription factor DNA sequence #578.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Pinus radiata.
XX
XX WO2000053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US006112.
XX
XX 11-MAR-1999; 99US-00266513.
XX
PR 18-AUG-1999; 99US-0149485P.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
XX WPI; 2000-579369/54.
XX
PT New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide.
XX
XX Claim 1; Page 593; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and
CC MYB
XX
SQ Sequence 316 BP; 100 A; 61 C; 84 G; 71 T; 0 U; 0 Other;
Query Match 6.6%; Score 123.6; DB 3; Length 316;
Best Local Similarity 68.4%; Pred. No. 1.5e-20;
Matches 171; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 1386 TCTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACTCAACGAGGTTGCCCTGA 1445
Db 7 TCTCCAACAATTAGCATGATTTCAGCAGCATGCTTGGAGGCCACAGAGAGACTTCCCGA 66
QY 1446 GAAATCTGTTTCGGTCTTACGGAATGGATGTTCCAAAACCTCTTCACTTACCCGAA 1505
Db 67 GCGATCTGTTTCTGCTTACGGGCTTGGCTATTGGAACATTTCTTCCTCGTATCCAAA 126
QY 1506 AGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAAGAGTCAGGTATCAAA 1565
Db 127 AGATCGACACACATATGCTCGGAGACACACTGGGCTTACAGAAATCAGGTCTCAA 186
QY 1566 CTGGTTTAAATGCGGGGTAGCTATGGAACCGCATGATGAGAGATGATGCGGA 1625
Db 187 TTGGTTTAAATGACGCTGATGACGCTTGGCTATGGAACCGCATGATGAGAGATGATGGA 246
QY 1626 AATGAACAAG 1635
Db 247 GGAACAAG 256

```

```

RESULT 8
AAC56228
ID AAC56228 standard; DNA; 462 BP.
XX
XX AAC56228;
AC
XX
XX 25-JAN-2001 (first entry)
DT
XX
DE Pinus radiata transcription factor DNA sequence #28.
XX
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
XX poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
XX basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
XX homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;
XX type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Pinus radiata.
XX
XX WO2000053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US006112.
XX
XX 11-MAR-1999; 99US-00266513.
XX
PR 18-AUG-1999; 99US-0149485P.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
XX WPI; 2000-579369/54.
XX
PT New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide.
XX
XX Claim 1; Page 138; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and
CC MYB
XX
SQ Sequence 462 BP; 154 A; 85 C; 124 G; 99 T; 0 U; 0 Other;
Query Match 6.6%; Score 123.6; DB 3; Length 462;
Best Local Similarity 68.4%; Pred. No. 1.7e-20;
Matches 171; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 1386 TCTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACTCAACGAGGTTGCCCTGA 1445
Db 7 TCTCCAACAATTAGCATGATTTCAGCAGCATGCTTGGAGGCCACAGAGAGACTTCCCGA 66
QY 1446 GAAATCTGTTTCGGTCTTACGGAATGGATGTTCCAAAACCTCTTCACTTACCCGAA 1505
Db 67 GCGATCTGTTTCTGCTTACGGGCTTGGCTATTGGAACATTTCTTCCTCGTATCCAAA 126
QY 1506 AGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAAGAGTCAGGTATCAAA 1565
Db 127 AGATCGACACACATATGCTCGGAGACACACTGGGCTTACAGAAATCAGGTCTCAA 186
QY 1566 CTGGTTTAAATGCGGGGTAGCTATGGAACCGCATGATGAGAGATGATGCGGA 1625

```


QY 1650 CA 1651
Db 1386 CA 1387

RESULT 11
ADD30303
ID ADD30303 standard; cDNA; 2385 BP.
XX
AC ADD30303;
XX
DT 15-JAN-2004 (first entry)
XX
Plant yield-related polynucleotide clone G1589.
XX
DE ds; transcription factor; transgenic plant; growth rate; senescence;
XX
KW seed germination rate; plant vigor; seedling vigor.
XX
OS Arabidopsis thaliana.
XX
PN WO2003013227-A2.
XX
PD 20-FEB-2003.
XX
PF 09-AUG-2002; 2002WO-US025805.
XX
PR 09-AUG-2001; 2001US-0310847P.
XX
PR 19-NOV-2001; 2001US-0336049P.
XX
PR 11-DEC-2001; 2001US-0336922P.
XX
PR 14-JUN-2002; 2002US-00171468.
XX
PA (MEND-) MENDEL BIOTECHNOLOGY INC.
XX
PI Ratcliffe O, Riechmann JL, Adam LJ, Dubell AT, Heard JE;
PI Pilgrim ML, Jiang C, Reuber TL, Creelman RA, Pineda O, Yu G;
PI Brown PB;
XX
WPI; 2003-248221/24.
DR P-PSDB; ADD30304.
XX
XX New plant transcription factor polynucleotides and polypeptides, useful
PT in producing transgenic plants with commercially valuable properties,
PT such as an alteration in a plant growth characteristic, e.g. growth rate
PT or apomixis.
XX
PS Disclosure; SEQ ID NO 332; 454pp; English.
XX
CC The invention relates to a number of isolated Arabidopsis thaliana cDNA
CC sequences and their encoded proteins which are especially transcription
CC factor related cDNA's and proteins. The isolated or recombinant plant
CC transcription factor polynucleotides and polypeptides are useful in
CC producing transgenic plants with commercially valuable properties, i.e.
CC modified or altered desirable traits as compared to a reference plant,
CC such as an alteration in a plant growth characteristic, e.g. growth rate,
CC germination rate of seeds, vigor of plants and seedlings, or leaf and
CC flower senescence. Sequence information related to the polynucleotides
CC and polypeptides can also be used in bioinformatic search methods. The
CC transgenic plant is useful for growing a progeny plant from a parent
CC plant. This sequence represents one of the cDNAs of the invention.
XX
SQ Sequence 2385 BP; 748 A; 578 C; 565 G; 494 T; 0 U; 0 Other;
Query Match 6.6%; Score 123.6; DB 9; Length 2385;
Best Local Similarity 69.4%; Pred. No. 3e-20; Indels 0; Gaps 0;
Matches 168; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 1410 CCATCAGATTTGGAGACCTCAACGAGTTTCCTCCAGAAATCTGTTTCGGTTCTACGGAA 1469
Db 1324 CAATAATGCTTTGGAGACCTCAACGAGTTTCCTCCAGAACGAGCCGCTTCAGTTCTCCGTCG 1383
QY 1470 TTGGATGTTCCAAACTTCTTCCCTTACCCCTTACCCGAAAGATTCGAGAAATCTTCTAGC 1529

1384 TTGGCTTTCGAACACTTTTTCATCATACCTCAAGGATTCGGACAGCATGCTAGC 1443
QY 1530 TATACGAAGTGGCTTGACAAGAGTCAAGTATCAAACTGGTTTATAAATGCGCGGTTAG 1589
Db 1444 TAAGCAAAACAGGACTCACTCGTAGCCAGGTGTGCAACTGTTTATAAACGCGAGATTGC 1503
QY 1590 GCTATGGAAGCCGATGATAGAGAGATGTATCGGAAATGAACAAGAGGAAGCTCAATAA 1649
Db 1504 GTTATGGAACCAATGTTGAGGAGATGTACATGAGGAAATCAAGCAGCAGCAAGAA 1563
QY 1650 CA 1651
Db 1564 CA 1565

RESULT 12
AAC55946
ID AAC55946 standard; DNA; 529 BP.
XX
AC AAC55946;
XX
DT 25-JAN-2001 (first entry)
XX
Rucalypus grandis transcription factor DNA sequence #77.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Eucalyptus grandis.
XX
XX WO2000053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US006112.
XX
XX 11-MAR-1999; 99US-00266513.
XX
PR 18-AUG-1999; 99US-0149485P.
XX
PR (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Shenk MA, Glenn M;
PI WPI; 2000-579369/54.
XX
XX New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide.
XX
PS Claim 1; Page 64-65; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and
CC MYB
XX
SQ Sequence 529 BP; 152 A; 127 C; 138 G; 112 T; 0 U; 0 Other;
Query Match 6.5%; Score 122.2; DB 3; Length 529;
Best Local Similarity 66.5%; Pred. No. 3.9e-20; Indels 0; Gaps 0;
Matches 175; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

1368 GTTCCACGAGCATTCCTCTTCCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACC 1427
52 GCTCAGGAGCAACAGAGCCCTACAGAGCTTGGGATGATGCAACAGCATGCATGAGGCC 111
1428 TCAACGAGGTTGCTGAGAAATCTGTTTCGTTCTACGGAATTCGGAATGTTCCAAACTT 1487
112 GCAAGAGAGGACTCTCTGAGAGTCTGTTTCTATTCTTCGGGCTTGGCTATTGAGCAFTT 171
1488 CTTTACCCCTTACCCGAAAGATTGCGAGAAACATCTTCTAGTATACGAAGTGGCTTGAC 1547
172 TCTTATCTCTTACCAAGGATCTGACAAATCTCTTTCGAAGGAGACAGGCTTGAC 231
1548 AAGAAGTCAGGATCAAACTGTTTATAATGCGCGGTTAGGCTATCGAAGCCGATGAT 1607
232 AAGAAGTCAGGTCCTGCAATGTTGTTTCAATCAATGCAAGAGTGGCTCTCTCGAAACCTATGGT 291
1608 AGAAGAGATGTATGCGGAATCA 1630
292 CGAAGAAATGTACAAAGAGAGA 314

RESULT 13
ABZ14349
ID ABZ14349 standard; DNA; 1599 BP.
XX
AC ABZ14349;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 2154.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US026685.
XX
PR 24-AUG-2000; 2000US-0227866P.
PR 26-JAN-2001; 2001US-0264647P.
PR 22-JUN-2001; 2001US-0300111P.
XX
PA (SCRI) SCRIPPS RES INST.
XX
PI (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Kreps J, Wang X, Zhu T;
XX
DR WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed and
PT producing plants with increased tolerance to these abiotic stresses.
XX
PS Claim 144; SEQ ID NO 2154; 577bp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising: (a) contacting nucleic acid
CC representative of expressed polynucleotides in the plant cell with an
CC array or probes representative of the plant cell genome; and (b)
CC detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention. Note: The sequence data for this patent is
CC not represented in the printed specification but is based on sequence
CC information supplied to Derwent by the European Patent Office
XX
SQ Sequence 1599 BP; 481 A; 329 C; 405 G; 384 T; 0 U; 0 Other;

Query Match 6.5%; Score 121.8; DB 6; Length 1599;
Best Local Similarity 71.9%; Pred. No. 7.3e-20;
Matches 159; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 1405 AAGAACCATCAGATTGGAGACCTCAACGAGGTTTCCTCGAGAAATCTGTTTCGGTTCTA 1464
Db 928 ATGCAACCTCAAGCTTGGAGACCTCAACGGGCTCTACCTCGAAACATCTGTTTGAATCTC 987
QY 1465 CGGAATTTGATGTTCACAAACTTCTTCACTCCCTTACCCGAAAGATTCTGATAAGATCATG 1524
Db 988 CGTCTTGGCTCTTTGAGCAATTTTCTTCACTTACCCCTTACCCAAAGGATTCTGATAAGATCATG 1047
QY 1525 CTAGCTATACGAAGTGGCTTGCACGAAGTCAAGTATCAAGATCTGTTTATAAATGCGCGG 1584
Db 1048 CTAGCAAGACAAACGGGCTTTGAGCCGGTTCAGGTTTCAAACTGGTTCATAAATGACGA 1107
QY 1585 GTTAGGCTATGGAAGCCGATGATAGAGATCATGTCGGA 1625
Db 1108 GTGCTTTATGGAAGCAATGGTGGAGGAGATATACAAGGA 1148

RESULT 14
AAC34112
ID AAC34112 standard; DNA; 833 BP.
XX
AC AAC34112;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 5490.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 22-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126284P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134219P.
PR 18-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137502P.
 PR 07-JUN-1999; 99US-0137724P.
 PR 08-JUN-1999; 99US-0138094P.
 PR 10-JUN-1999; 99US-0138540P.
 PR 14-JUN-1999; 99US-0138847P.
 PR 16-JUN-1999; 99US-0139119P.
 PR 16-JUN-1999; 99US-0139452P.
 PR 17-JUN-1999; 99US-0139453P.
 PR 17-JUN-1999; 99US-0139492P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139763P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 24-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140695P.
 PR 28-JUN-1999; 99US-0140823P.
 PR 29-JUN-1999; 99US-0140991P.
 PR 30-JUN-1999; 99US-0141287P.
 PR 01-JUL-1999; 99US-0141842P.
 PR 01-JUL-1999; 99US-0142154P.
 PR 02-JUL-1999; 99US-0142055P.
 PR 06-JUL-1999; 99US-0142390P.
 PR 08-JUL-1999; 99US-0142803P.
 PR 09-JUL-1999; 99US-0142920P.
 PR 12-JUL-1999; 99US-0142977P.
 PR 13-JUL-1999; 99US-0143542P.
 PR 14-JUL-1999; 99US-0143624P.
 PR 15-JUL-1999; 99US-0144005P.
 PR 16-JUL-1999; 99US-0144083P.
 PR 16-JUL-1999; 99US-0144086P.
 PR 19-JUL-1999; 99US-0144321P.
 PR 19-JUL-1999; 99US-0144331P.
 PR 19-JUL-1999; 99US-0144332P.
 PR 19-JUL-1999; 99US-0144333P.
 PR 19-JUL-1999; 99US-0144334P.
 PR 19-JUL-1999; 99US-0144335P.
 PR 20-JUL-1999; 99US-0144352P.
 PR 20-JUL-1999; 99US-0144632P.
 PR 20-JUL-1999; 99US-0144884P.
 PR 21-JUL-1999; 99US-0144814P.
 PR 21-JUL-1999; 99US-0145086P.
 PR 21-JUL-1999; 99US-0145088P.
 PR 22-JUL-1999; 99US-0145089P.
 PR 22-JUL-1999; 99US-0145087P.
 PR 22-JUL-1999; 99US-0145089P.
 PR 22-JUL-1999; 99US-0145192P.
 PR 23-JUL-1999; 99US-0145145P.
 PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145222P.
 PR 26-JUL-1999; 99US-0145276P.
 PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.

PR 02-AUG-1999; 99US-0146388P.
 PR 02-AUG-1999; 99US-0146389P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 04-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 05-AUG-1999; 99US-0147360P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 06-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147935P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.
 PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 26-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0155659P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-015753P.
 PR 05-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159285P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.

PR	28-OCT-1999;	99US-0161920P.		PR	14-MAY-1999;	99US-0134221P.
PR	28-OCT-1999;	99US-0161992P.		PR	14-MAY-1999;	99US-0134370P.
PR	28-OCT-1999;	99US-0161993P.		PR	18-MAY-1999;	99US-0134768P.
PR	29-OCT-1999;	99US-0162142P.		PR	19-MAY-1999;	99US-0134941P.
<hr/>						
Query Match 6.4%; Score 121.2; DB 3; Length 833;						
Best Local Similarity 69.3%; Pred. No. 8.1e-20;						
Matches 165; Conservative 0; Mismatches 73; Indels 0; Gaps 0;						
<hr/>						
QY	1388	TTCAGCAGCTGAACAAAGAACCATCACGATTGGAGACTCAACGAGGTTCCTCGTGA	1447	PR	27-NAY-1999;	99US-0136392P.
Dd	105	TTCCAGCGGTAGTGTFGATGCACACCTCACACITGGCGGCTCAACCGCGTTTTACCA	164	PR	28-MAY-1999;	99US-0136782P.
QY	1448	AATCTGTTTCGGTTCTACGGAATTGGATGTTCCAAAACCTTCCTTCACCCCTTACCCGAA	1507	PR	01-JUN-1999;	99US-0137222P.
Dd	165	CTTCTGTTTGGTTCTCCGTTGCTATTGAGCATTTCTCCACCCCTTATCCAAAG	224	PR	03-JUN-1999;	99US-0137528P.
QY	1508	ATTCCGAGAACATCTTCTAGCTATACGAAGTGGCTTTGACAAGTTCAGGTATCAAAC	1567	PR	04-JUN-1999;	99US-0137502P.
Dd	225	ATTACACAAAGATCATGCTCGCTAGACAAACGGGTTGAGCCGAGGCCAGGTATCGA	284	PR	07-JUN-1999;	99US-0137724P.
QY	1568	GCTTTATAATGCGCGGTTAGCTATGGAAGCCGATGATAGAGATGTTATGCGGA	1625	PR	08-JUN-1999;	99US-0138094P.
Dd	285	GCTTCATAATGCGGTGTGCGTCTCTGGAACCGGATGTTGGAGGAGATGTACAAG	342	PR	10-JUN-1999;	99US-0138540P.
<hr/>						
RESULT 15						
ID	AAC43406			PR	10-JUN-1999;	99US-0138847P.
XX	AAC43406 standard; DNA; 1449 BP.			PR	14-JUN-1999;	99US-0139119P.
AC	AAC43406;			PR	16-JUN-1999;	99US-0139452P.
XX				PR	17-JUN-1999;	99US-0139453P.
DT				PR	18-JUN-1999;	99US-0139454P.
DE	17-OCT-2000 (first entry)			PR	18-JUN-1999;	99US-0139456P.
XX				PR	18-JUN-1999;	99US-0139457P.
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 39140.			PR	18-JUN-1999;	99US-0139458P.
XX				PR	18-JUN-1999;	99US-0139459P.
XX				PR	18-JUN-1999;	99US-0139460P.
KW	Hybridisation assay; genetic mapping; gene expression control;			PR	18-JUN-1999;	99US-0139461P.
KW	protein identification; signal transduction pathway; metabolic pathway;			PR	18-JUN-1999;	99US-0139462P.
KW	promoter; termination sequence; ss.			PR	18-JUN-1999;	99US-0139463P.
XX				PR	18-JUN-1999;	99US-0139464P.
OS	Arabidopsis thaliana.			PR	18-JUN-1999;	99US-0139465P.
PN				PR	18-JUN-1999;	99US-0139466P.
PD	EP1033405-A2.			PR	18-JUN-1999;	99US-0139467P.
XX				PR	18-JUN-1999;	99US-0139468P.
PF	06-SEP-2000.			PR	18-JUN-1999;	99US-0139469P.
XX				PR	18-JUN-1999;	99US-0139470P.
XX	25-FEB-2000; 2000BP-00301439.			PR	18-JUN-1999;	99US-0139471P.
PR				PR	18-JUN-1999;	99US-0139472P.
PR	25-FEB-1999;	99US-0121825P.		PR	18-JUN-1999;	99US-0139473P.
PR	05-MAR-1999;	99US-0123180P.		PR	18-JUN-1999;	99US-0139474P.
PR	09-MAR-1999;	99US-0123548P.		PR	18-JUN-1999;	99US-0139475P.
PR	23-MAR-1999;	99US-0125788P.		PR	18-JUN-1999;	99US-0139476P.
PR	25-MAR-1999;	99US-0126264P.		PR	18-JUN-1999;	99US-0139477P.
PR	29-MAR-1999;	99US-0126785P.		PR	18-JUN-1999;	99US-0139478P.
PR	01-APR-1999;	99US-0127462P.		PR	18-JUN-1999;	99US-0139479P.
PR	06-APR-1999;	99US-0128234P.		PR	18-JUN-1999;	99US-0139480P.
PR	08-APR-1999;	99US-0128714P.		PR	18-JUN-1999;	99US-0139481P.
PR	16-APR-1999;	99US-0129845P.		PR	18-JUN-1999;	99US-0139482P.
PR	19-APR-1999;	99US-0130077P.		PR	18-JUN-1999;	99US-0139483P.

```
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151068P.
PR 27-AUG-1999; 99US-0151088P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155653P.
PR 28-SEP-1999; 99US-0158458P.
PR 29-SEP-1999; 99US-0158596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.

PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 6.4%; Score 119.8; DB 3; Length 1449;
Best Local Similarity 87.3%; Pred.No. 2.2e-19;
Matches 169; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 1375 CAGCATTCCTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTTGGAGACCTCAACGA 1434
Db |||||
Qy 811 CAACAAAGAGCGTTGCAGCGTTAGGTGTGATGCAACCTCACACTTGGCGGCTCAACGC 870
Db |||||
Qy 1435 GGTTCGCTGAGAAATCTGTTTCGGTTCTACGGAATGGATGTTCCAAAACCTTCCTTCAC 1494
Db |||||
Qy 871 GGTTCACAGATTCCTTCTGTTTGGTTCTCCGGTCTTGCTATTGAGCATTTCTCCAC 930
Db |||||
Qy 1495 CCTTACCCGAAAGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAGAAGT 1554
Db |||||
Qy 931 CCTTATCCAAAGATTTCAGACAAGATCATGCTCGCTAGACAAACGGGTTGAGCCGAGGC 990
Db |||||
Qy 1555 CAGGTATCAAACTGGTTTATAAATGCGGGTTAGGCTATGGAAGCCGATGATAGAGAG 1614
Db |||||
Qy 991 CAGGTATCGAACTGGTTTCATAAATGCGGCTGTGCGTCTCTGGAACCCGATGCTGGAGGAG 1050
Db |||||
Qy 1615 ATGTATGCGGA 1625
Db |||||
Qy 1051 ATGTACAAGA 1061
```

Search completed: September 2, 2004, 06:19:12
Job time : 769 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 2, 2004, 04:48:37 ; Search time 142 Seconds
(without alignments)
7370.686 Million cell updates/sec

Title: US-09-423-575-1
Perfect score: 1886
Sequence: 1 atttagttataaaattgtgc.....ttaaaaaaaaaaaaaaaa 1886

Scoring table: IDENTITY NUC
Gap0 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A-COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B-COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS-COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	2.5	2305	US-09-976-594-1031	Sequence 1031, Ap
2	47.6	2.5	1248	US-09-489-847-101	Sequence 101, App
3	45.2	2.4	1579	US-09-620-312D-927	Sequence 927, App
C 4	44.4	2.4	780	US-09-134-000C-2237	Sequence 2237, Ap
5	44.4	2.4	832	US-09-621-976-2813	Sequence 2813, Ap
C 6	42	2.2	7218	US-08-232-463-14	Sequence 14, Appl
7	41	2.2	2418	US-09-601-198-61	Sequence 61, Appl
8	40.8	2.2	1770	US-09-976-594-195	Sequence 195, App
9	40	2.1	998	US-09-122-400B-5	Sequence 5, Appl
10	40	2.1	7218	US-08-232-463-14	Sequence 14, Appl
11	39.8	2.1	1364	US-09-428-034-1	Sequence 1, Appl
12	39.2	2.1	535	US-09-621-976-683	Sequence 683, App
C 13	38.6	2.0	6113	US-10-204-708-13	Sequence 13, Appl
14	38.4	2.0	2049	US-09-624-945-2	Sequence 2, Appl
C 15	38.4	2.0	2110	US-09-419-459-1	Sequence 1, Appl
16	38.4	2.0	2952	US-08-306-691B-54	Sequence 54, Appl
C 17	38.2	2.0	6317	US-10-204-708-11	Sequence 11, Appl
18	38.2	2.0	9098	US-09-358-082A-28	Sequence 28, Appl
C 19	38	2.0	1398	US-08-956-171B-662	Sequence 662, App
C 20	38	2.0	7424	US-09-596-002-8	Sequence 8, Appl
21	37.8	2.0	816	US-09-134-000C-2630	Sequence 2630, Ap
C 22	37.6	2.0	6609	US-09-976-594-690	Sequence 690, App
23	37.6	2.0	9636	US-08-323-170B-1	Sequence 1, Appl
C 24	37.6	2.0	9636	US-08-954-441-1	Sequence 1, Appl
25	37.4	2.0	1016	US-09-716-129-42	Sequence 42, Appl
C 26	37.2	2.0	350	US-09-598-401C-23	Sequence 23, Appl
C 27	37.2	2.0	832	US-09-621-976-2813	Sequence 2813, Ap

28	37.2	2.0	960	4	US-09-134-001C-217	Sequence 217, App
29	37	2.0	1720	3	US-08-655-352-11	Sequence 11, Appl
30	37	2.0	1720	4	US-09-258-016-11	Sequence 11, Appl
31	37	2.0	1720	4	US-09-257-825B-11	Sequence 11, Appl
C 32	36.8	2.0	10640	4	US-09-417-485D-5	Sequence 5, Appl
33	36.8	2.0	1892	4	US-09-665-189A-67	Sequence 67, Appl
C 34	36.6	1.9	640681	4	US-09-790-988-1	Sequence 1, Appl
35	36.4	1.9	8878	1	US-08-206-176-3	Sequence 3, Appl
36	36.4	1.9	11015	4	US-10-204-708-55	Sequence 55, Appl
C 37	36.4	1.9	53332	4	US-09-801-861-3	Sequence 3, Appl
38	36.2	1.9	1565	4	US-08-956-171B-527	Sequence 527, App
C 39	36.2	1.9	5340	4	US-09-627-122-21	Sequence 21, Appl
C 40	36.2	1.9	6326	4	US-10-204-708-57	Sequence 57, Appl
41	36	1.9	1057	4	US-08-956-171B-946	Sequence 946, App
C 42	36	1.9	1078	4	US-08-956-171B-661	Sequence 661, App
43	36	1.9	2079	3	US-09-381-849-4	Sequence 4, Appl
C 44	36	1.9	6801	4	US-10-204-708-62	Sequence 62, Appl
C 45	35.8	1.9	207	4	US-09-543-681A-2737	Sequence 2737, Ap

ALIGNMENTS

RESULT 1
US-09-976-594-1031
; Sequence 1031, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1443
; SOFTWARE: PERL Program
; SEQ ID NO 1031
; LENGTH: 2905
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 480885.2
US-09-976-594-1031

Query Match	2.5%	Score 48;	DB 4;	Length 2905;
Best Local Similarity	52.5%	Pred. No. 0.0056;		
Matches 105;	Conservative	0;	Mismatches 95;	Indels 0;
Gaps	0;			
QY	1471	TGGATGTTCCAAACTTCCTTACCCCTTACCCGAAAGATTCCGAGAAACATCTTCTAGCT	1530	
Db	1130	TGGCTCTTCCAGCATCTCACACATCCGTACCTTCCGAGAGCAGAAACAGTTAGCG	1189	
QY	1531	ATACGAAGTGCTTGACAAGAGTCAGGTATCAACTGGTTTATAATGCGCGGTAGG	1590	
Db	1190	CAAGACAGAGACTTCAATCTCCAGTAAGAACTGGTTTATTAATGCCAAGAAGA	1249	
QY	1591	CTATGGAAGCCGATGATAGAGAGATGTATCGGAAATGAACAGAGAGCTCAATAAC	1650	
Db	1250	ATAGTACAGCCCATGATTGACCAAGTCAATCGAGCAGTGGAGCCAGGAGCATATAGT	1309	
QY	1651	AGTCACATTCAACCCACGG	1670	
Db	1310	CCAGAGGTCAGCCCATGGG	1329	

RESULT 2
US-09-489-847-101
; Sequence 101, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:

```

; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1998-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 101
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-489-847-101

Query Match      2.5%; Score 47.6; DB 4; Length 1248;
Best Local Similarity 55.4%; Pred. No. 0.0048;
Matches 92; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 163 AATAAGCATTATCTCTCTTCAGTTTTTAAACACACATTTTGGAAATTTTGATGTAAAAA 222
      |||||
Db 1067 AATACGCTAGATTTTGTCTCTCTGTATGTGACACACATTTTGTACATTATGTTATTGGAA 1126

Qy 223 TTCTCTTTTGGACGTTCTGTGTCTGAAATCTTCCCAAGGTTCTTATCAGAAGAGGAGG 282
      |||||
Db 1127 TTTCCTTCATACATTATTTCTCTTAAACTCTCAAAAAAATAAAAAAAAAAAAAA 1186

Qy 283 ATAAAGTTTTCATAGAAACCCCAATGGACAACACAAACACAAACAACA 328
      |||||
Db 1187 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1232

RESULT 3
US-09-620-312D-927
; Sequence 927, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/489,725
; PRIOR FILING DATE: 2000-01-21

```



```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
; US-08-232-463-14

Query Match 2.2%; Score 42; DB 1; Length 7218;
Best Local Similarity 7.5%; Pred.No. 0.36;
Matches 33; Conservative 210; Mismatches 195; Indels 0; Gaps 0

QY 1287 GAGAGAGAGATCTGCAAGAAGATATCTCTATGGATCTGTATGGAGAGAGGCAAGA 1346
DB 1431 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1372
QY 1347 CAAGACTCAAGAAACCTCTATGTTCCACGACGATTCCTCTTCAGCAGCTCAACGAA 1406
DB 1371 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1312
QY 1407 GAACATCAGATTGGAGACCTCAACGAGGTTGCTGAGAAATCTGTTCCGTTCTACG 1466
DB 1311 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1252
QY 1467 GAATTGGATGTTCCAAATCTCCTTACCCGTACCCGAAAGATTCGAGAAACATCTTCT 1526
DB 1251 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1192
QY 1527 AGCTATCAGAAAGTGGCTTGACAGAAGTCAGTATCAAACTGGTTTAAATGCGCGGT 1586
DB 1191 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1132
QY 1587 TAGCCTATGAAGCGGATGATAGAGAGATGTATCGGAATGAACAAGAGAGACTCAA 1646
DB 1131 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1072
QY 1647 TAACAGTCACATTCAACCAACGACCAACTCTTCGAATGCCAAATCTGTATGATGAG 1706
DB 1071 RRRRATCGAAGTCCCTCGACCTGCACCAAGCTCGGAATTAATCTGTGACGATG 1012
QY 1707 CCAAGCAATGCATAATA 1724
DB 1011 GCAAAACGAAGAAAATA 994

RESULT 7
US-09-601-198-61
; Sequence 61, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.

```

RESULT 7
US-09-601-198-61
; Sequence 61, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.

```

; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match      2.4%; Score 44.4; DB 4; Length 832;
Best Local Similarity 13.4%; Pred.No.0.029;
Matches 50; Conservative 168; Mismatches 149; Indels 7; Gaps 17

QY      570 TCCTCATCTCTCCATCTTCATCTCTGGATCATTTTAAGACACATGATGATCTCTCAAA 629
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Ddb      11 WYACWTWKWSWSYMYWKYMYKTYWRWRKKKKAHWKYKWTWTWYWRVAMGTYYK 70
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      630 CAACATGTGGGGTTTTGAAGCAAACTAGTGTTCAGGCATTTTCAGGTGTAGTTGGTCC 689
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Ddb      71 KAMCRTRTKKKKKGGYMMWYWGWRSSYMAWTRTWTGYAYRSMMYWRVRCWKKAYY 130
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      690 AAGTGAACCAATGATGCTCATCTCGGTGAAGAGATTTCCC-----GTTTCTAAATT 742
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Ddb      131 RKTTCYSGKGTWWRWKKKATTTWKKTYTAAATRYMMCMWKWRASWYCWNGKA 190
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      743 CGAATAAAGAAACAATGAGCTTTCATGTAGTCTTCGATCAGAGTGTTCGTGAATGCT 802
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Ddb      191 RKMSTWRKRSYASAGSAKCCYSCSNGAMSWKMYWRWRGWATGAGMKAWRASCMNR 250
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      803 CGGAGATAAGTCTTTGTGCAGCTACAAGATTAGCCTCAGAGCAAGCTTCTTCGACGACGA 862
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Ddb      251 RYAGKSKTSYKSMWMCWTRSWKYCYTKARWTGYCYRGGMMGKRGRTYASKYNNKRW 310
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      863 AAGACATTTCTAAATAAGTGTGTACTCAAGGTTTCTCAAGTTTCTCAATATATTGGCTCAAAAT 922
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Ddb      311 WWCWARYRYSTGTFRASMNWRERYTMMKWKYAWABAARWAWMWAWRACAAATA 370
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      923 ACCTTCACCTGTT 936
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Ddb      371 TAATTATTATGGT 384
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F.G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

/ / STREET: 1800 Diagonal Road.
/ / CITY: Alexandria
/ / STATE: VA
/ / COUNTRY: USA
/ / ZIP: 22313-0299
/ /
/ / COMPUTER READABLE FORM:
/ / MEDIUM TYPE: Floppy disk
/ /
/ / COMPUTER: IBM PC compatible
/ /

```

; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; TITLE OF INVENTION: UREALYTICUM
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; PRIOR FILING DATE: 2000-12-08
; PRIOR FILING DATE: 2000-12-08
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 61
; LENGTH: 2418
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-61

Query Match 2.2%; Score 41; DB 4; Length 2418;
Best Local Similarity 44.3%; Pred. No. 0.39; Mismatches 210; Indels 0; Gaps 0;
Matches 167; Conservative 0;

QY 1 ATTAGTTATATAAAAGTTGCTATTTGTTGATCTAGTCTCTGAATCTTTTAGTGAGCGAG 60
DB |||||
QY 61 ATGATCAAGATTATGAAATTTCTTCATGAAATTTGTAAGAAAGACATAGAGAGCT 120
DB |||||
QY 202 TAGGTTAAATTTTATTGTCATCTGTTTAAATATGCAATGCTGTTTAAATGCTAAT 261
QY 121 GCGGAATGAAAGTACACTGTTCTTTCACGCGAAGAGATAAATAAGCATTATCTTCT 180
DB |||||
QY 262 TCTGTTTAAAGGATTAATTTGTTGATTAATTAATAAATTTGCTCATGCTGTTTATC 321
QY 181 CTTGAGTTTAAACACACATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 240
DB |||||
QY 322 AACAAATATTAACAATGATAGTGTGTTGATCTAAATCTTTTGTGTTAAATCTAT 381
QY 241 GTTGTCTGAAATCTTCCCAAGGTTCTATCAGAGAGAGATAAAGTTTTCATAGAAAC 300
DB |||||
QY 382 TTATCTTATTAAATCACTTTAGTAGGACAAATTAATCTTTTATTATTATCTCAAT 441
QY 301 CCAATGGAC 360
DB |||||
QY 442 AAAATGAGCAACAGCATCTTGAGTTGTTTATTATTAAATATCAATTAATGTCATAAT 501
QY 361 CAAAATCTCTCTCTCAT 377
DB |||||
QY 502 TAAATTTCTAAATCAT 518

RESULT 8
US-09-976-594-195
; Sequence 195, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 195
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID No. 6673549 995887.12
; NAME/KEY: unsure
; LOCATION: 846, 851
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-195
Query Match 2.2%; Score 40.8; DB 4; Length 1770;
Best Local Similarity 55.7%; Pred. No. 0.38; Mismatches 62; Indels 0; Gaps 0;
Matches 78; Conservative 0;

QY 1471 TGGATGTTCCAAATCTCTTACCCGAAAGATTTCGAGAAACATCTTCTAGCT 1530
DB |||||
QY 1029 TGGTGTTCAGCACCTCTCGCACCCGCTACCCCTCGGAGGAGCAGAAACAGCTGGG 1088
QY 1531 ATACGAAGTGGTTGACAAAGATCAGGTATCAAACTGTTTATAAATCGGGGTTAG 1590
DB |||||
QY 1089 CAGGACACGGGCTCACCATCTCTGCAAGTCAACAACTGTTTCAATTAACCCCGAGACGC 1148
QY 1591 CTATGGAAGCCGATGATAGA 1610
DB |||||
QY 1149 ATCGTCAACCTATGATCGA 1168

RESULT 9
US-09-122-400B-5
; Sequence 5, Application US/09122400B
; Patent No. 6245974
; GENERAL INFORMATION:
; APPLICANT: Michalowski, Susan
; APPLICANT: Spiker, Steven
; TITLE OF INVENTION: MATRIX ATTACHMENT REGIONS
; FILE REFERENCE: Michalowski and Spiker
; CURRENT APPLICATION NUMBER: US/09/122,400B
; CURRENT FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 60/066,118
; PRIOR FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 5
; LENGTH: 998
; TYPE: DNA
; ORGANISM: Nicotiana tabacum
US-09-122-400B-5

Query Match 2.1%; Score 40; DB 3; Length 998;
Best Local Similarity 47.0%; Pred. No. 0.48; Mismatches 124; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 147 ACGGAGAGAGAGATAAATAGCATATCTCTTCTGAACTGTGTCTGAAATCTTCCCAAGGTT 266
DB |||||
QY 420 ACTGAAAAAAGAGAGCTGAAATCAATTTCTAAAGCAATTTTATTTGTAATAATCTGA 479
QY 207 AATTTTGATGTAAAAATCTCTTTGAACTGTGTGTCTGAAATCTTCCCAAGGTT 266
DB |||||
QY 480 AAAAACTACTAAAAATCTGAAAAATGAAAAATATTTTTTTCTAATTTTACAAAAAAA 539
QY 267 TATCAGAAAGAGATAAAGTTTCATAGAAACCAATGGACACACACACACACACAC 326
DB |||||
QY 540 CTGCTTTAAAAAAGCTGAAAAATATTTCTTAAACAAATATTTTGTAAAAACATAAAAA 599
QY 327 CACTTTTAGTCTCTGGATAATGTCATGACTACCAAAATCTCTCTCATGGATTTAT 386
DB |||||
QY 600 AATATTTTCTTCTTTTCTGATTTTATTTAGTTAAAAATATTTAAAGTTTTCAGTTT 659
QY 387 ACCTTCAAGAGAGATCAATTC 410
DB |||||
QY 660 AATTACTTTAGAAAAATTAATTTTCT 683

RESULT 10
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367

```
/ GENERAL INFORMATION:
/ APPLICANT: DORNER, F.
/ APPLICANT: SCHEIFLINGER, F.
/ APPLICANT: FALKNER, F. G.
/ TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
/ NUMBER OF SEQUENCES: 52
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Foley & Lardner
/ STREET: 1800 Diagonal Road, Suite 500
/ CITY: Alexandria
/ STATE: VA
/ COUNTRY: USA
/ ZIP: 22313-0299
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/232,463
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/07/935,313
/ FILING DATE:
/ APPLICATION NUMBER: EP 91 114 300.6
/ FILING DATE: 26-AUG-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BENT, Stephen A.
/ REGISTRATION NUMBER: 29,768
/ REFERENCE/DOCKET NUMBER: 30472/114 IMMU
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703)836-9300
/ TELEFAX: (703)683-4109
/ TELEX: 899149
/ INFORMATION FOR SEQ ID NO: 14:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 7218 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ CLONE: PTZ9pt-F1s
/ US-08-232-463-14

Query Match.
Best Local Similarity 2.1%; Score 40; DB 1; Length 7218;
Matches 7; Conservative 187; Mismatches 132; Indels 0; Gaps 0;

QY 329 CTTTAGTTCTCTGGATAATGCTAGTAACTCAACAAATCTCTTCTCATGATTTTATAC 388
Db 1118 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1177
QY 389 CTTCAAGAAGATTCAACTTCTATCTCAACATGCTTCATGGAATACCATCAGATCAG 448
Db 1178 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1237
QY 449 ATCTCTCAAAATGGTGGCTTTGATTTTCAATCTATGCTGACTAACAAATCTTAT 508
Db 1238 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1297
QY 509 CATCTTCTCAGGCTATGATGATTCAGATAACCGAATGTTAGITCATGCTCTC 568
Db 1298 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1357
QY 569 CTCTCATCTCTCTCCACTTCATCTTGGATCATTTAAGACACTATGATTCCTCAA 628
Db 1358 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1417
QY 629 ACAACATGTGGGTTTGAAGCAAT 654
Db 1418 YYYYYYYYYYYYYYGTACCAAT 1443
```

```
RESULT 11
US-09-428-034-1
; Sequence 1, Application US/09428034
; Patent No. 6428996
; GENERAL INFORMATION:
; APPLICANT: Cheng, Kuo-Joan
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Teai, Cheng-Fang
; APPLICANT: Hsu, Yih-Chin
; TITLE OF INVENTION: CELLULOSE ENZYMES
; FILE REFERENCE: 08919/036001
; CURRENT APPLICATION NUMBER: US/09/428,034
; CURRENT FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1364
; TYPE: DNA
; ORGANISM: Piromyces rhizinflata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1128)
US-09-428-034-1

Query Match
Best Local Similarity 2.1%; Score 39.8; DB 4; Length 1364;
Matches 71; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 1764 AATTGCTACTCTATGATTGCCAAAACCTAAACCATGTAGCATATCATTACGTATGTTA 1823
Db 1110 AAATGCTAATCCCACTTCTAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1169
QY 1824 TAATTGTATATACAACTCCCTTTATCTTTGACTATTTTCATTTTAAAAAATAAATA 1883
Db 1170 TAAATATATAATTATTTTATTTTTCATTTACTATTAATAATAATAATAATAATA 1229
QY 1884 AAA 1886
Db 1230 TAA 1232

RESULT 12
US-09-621-976-683
; Sequence 683, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 683
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 109..534
US-09-621-976-683

Query Match
Best Local Similarity 2.1%; Score 39.2; DB 4; Length 535;
Matches 89; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 1412 ATCAGATTGGAGACCTCAACAGAGTTTGCCTGAGAAATCTGTTTCGTTCTACGGATT 1471
Db 251 AGCAAGAAGAGAGCGAAGGAACTTGCACCGAGTCGTTAAGATCCTCGCGACT 310
```

QY 1472 GGATGTTCCAAACTTCCTTCCCTTACCCGAAAGATTCCGAGAAACATCTTCTAGCTA 1531
| | | | |
Db 311 GGATGTTAAGCATCGGTTTAAGGCTACCTTTCAGAGAAGAGCAAAATGCTGTGAG 370
| | | | |
QY 1532 TACGAAGTGGCTTGACAGAAATCAGGTATCAAACTGGTTTATAAATCGCGG 1583
| | | | |
Db 371 AGAAGACCAATTTGCTTTGTTGCGGATTTCTTAAGTGGTTTATCAATGCTCG 422
| | | | |

RESULT 13

US-10-204-708-13/c
; Sequence 13, Application US/10204708
; Patent No. 667731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 13
; LENGTH: 6113
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-13

Query Match 2.0%; Score 38.6; DB 4; Length 6113;
Best Local Similarity 46.8%; Pred. No. 2.7;
Matches 155; Conservative 0; Mismatches 174; Indels 2; Gaps 1;
QY 189 TTTACACACATTTTGGAAATTTGATGTAAATCTCTTTGGAAACGTTGTGTCTG 249
| | | | |
Db 3637 TTTATCACTCAATTAATTTATTTTACCTAACCTTTTCTTATACTTATTTTAAATTTA 3578
| | | | |
QY 249 AAAT--CTTCCCAAGGTTCTATCAGAAGAAGAGGATAAAGTTTCATAGAAACCCCAATG 306
| | | | |
Db 3577 AACTTATTTCCCTATAAATCTCTATCAACAANAATATATCTCACAAAAAATAAATA 3518
| | | | |
QY 307 GACACACACACACACACACATTTTGTCTCTGGAATATGTCATGACTAACCAAAAT 366
| | | | |
Db 3517 AAAACCTCAAAAAACATAACATAATATAATATTTCTTAACAAATTAACAAACAAAAA 3458
| | | | |
QY 367 CCTCTTCTCATGGATTTTATACCTTCAAGAGAAGATTCAACTTCTTCAACATGCTT-426
| | | | |
Db 3457 ATATTTCTTTTCCATAAATTTATCAATAAAATTAATAATATCAAACTATCAAACTAAAT 3398
| | | | |
QY 427 CCATGGAATACATCAGATCAGATCCTCTACAAATGGGTGGCTTTGATATTTTCAATCT 486
| | | | |
Db 3397 TTTTAAATTAATAATATAATTTACATACAATAAATCAACCTTTTATAATACAAATTT 3338
| | | | |
QY 487 ATGCTGACTCAAAATTAATCTATCTCTC 517
| | | | |
Db 3337 ATTATTTTAAATAAATAAATAAACCTC 3307
| | | | |

RESULT 14

US-09-624-945-2
; Sequence 2, Application US/09624945

; Patent No. 6607915
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Wanciewicz, Edward
; TITLE OF INVENTION: Antisense Modulation of E2A-Pbx1 Expression
; FILE REFERENCE: ISPH-0477
; CURRENT APPLICATION NUMBER: US/09/624,945
; CURRENT FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/156,836
; PRIOR FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2049
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-624-945-2

Query Match 2.0%; Score 38.4; DB 4; Length 2049;
Best Local Similarity 58.9%; Pred. No. 1.8;
Matches 66; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
QY 1493 ACCCTTACCCGAAAGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTCAAGAA 1552
| | | | |
Db 1133 ACCCTTACCCGAAAGATTCCGAGAAAGAGGAGTTAGCCAAAGAGTGGCATCACAGTCT 1192
| | | | |
QY 1553 GTCAGGTATCAAACTGGTTTATAAATCGCGGGTTAGGCTATGGAAGCCGAT 1604
| | | | |
Db 1193 CCCAGGTATCAAACTGGTTTGAATAAAGCAATCCGTTACAAAGAAACAT 1244
| | | | |

RESULT 15

US-09-419-459-1/c
; Sequence 1, Application US/09419459
; Patent No. 6222028
; GENERAL INFORMATION:
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Cheng, Kuo-Juan
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Chang, Chia-Chieh
; TITLE OF INVENTION: CELLULOSE ENZYMES
; FILE REFERENCE: 08919/037001
; CURRENT APPLICATION NUMBER: US/09/419,459
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2110
; TYPE: DNA
; ORGANISM: Piromyces rhizinflata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1749)
US-09-419-459-1

Query Match 2.0%; Score 38.4; DB 3; Length 2110;
Best Local Similarity 49.0%; Pred. No. 1.9;
Matches 102; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 1679 TTCGAATGCCAAATCTGTTATGATGAGCCAAAGCAATGCAATAAATAGCAACAATTTGTG 1738
| | | | |
Db 2062 TTTAAATATTTTACACAATTTTATAATCTTATTAGCTATTATTATTAATATACATATCCA 2003
| | | | |
QY 1739 TTTACCACTTTGTGATAATTAGGCAATTTGCTATGATGCCCCAAAACCTAAACCA 1798
| | | | |
Db 2002 TCTTTAAATAATCTTGATTTTAATTAATCTTTTTTTGCTTTAAGTAAAGATAAAT 1943
| | | | |
QY 1799 TGTAGACTATCATTCAGTATGTTATAATGTATATACTCCCTTTATCTTTGACTATT 1858
| | | | |
Db 1942 AATTAAATTTTATTTTCATTCATTATTATTATTTTATTTTCTTTCTTATAAATTT 1883
| | | | |
QY 1859 TCATTTTATTAATAAATAAATAAATAAATAA 1886
| | | | |

Db 1882 TTTTATAAATAATAAGAAATAAAA 1855

Search completed: September 2, 2004, 08:25:14
Job time : 144 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.
OM nucleic - nucleic search, using sw model
Run on: September 2, 2004, 08:22:57 ; Search time 887 Seconds
(without alignments)
10568.913 Million cell updates/sec

Title: US-09-423-575-1
Perfect score: 1886
Sequence: 1 attatgtataaaattgtgc.....ttaaaaaaaataaaaaaa 1886

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3267054 seqs, 2485319735 residues

Total number of hits satisfying chosen parameters: 6534108

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171.6	9.1	422	13	US-10-424-599-50272
2	130.2	6.9	1555	13	US-10-425-114-13353
3	130.2	6.9	1995	17	US-10-437-963-7710
4	130.2	6.9	3033	13	US-10-424-599-130628
5	126.4	6.7	2799	17	US-10-437-963-31585
6	126.2	6.7	592	17	US-10-021-323-14220
7	125.2	6.6	1716	13	US-10-425-114-9882
8	124.4	6.6	1452	13	US-10-425-114-5047
9	124.4	6.6	1606	13	US-10-425-114-7804
10	124.2	6.6	1785	17	US-10-437-963-19673
11	123.6	6.6	2043	9	US-09-938-842A-1337
12	123.6	6.6	2043	11	US-09-938-842A-1337
13	123.6	6.6	2385	13	US-10-225-066A-335
14	123.6	6.6	2385	16	US-10-374-780A-2671

15	122.6	6.5	227	12	US-09-922-293-2000	Sequence 2000, Ap
16	121.8	6.5	1599	9	US-09-938-842A-2154	Sequence 2154, Ap
17	121.8	6.5	1599	11	US-09-938-842A-2154	Sequence 2154, Ap
18	121.4	6.4	1280	13	US-10-425-114-13349	Sequence 13349, A
19	121.4	6.4	1822	13	US-10-424-599-133558	Sequence 133558, A
20	119.2	6.3	3029	17	US-10-437-963-60154	Sequence 60154, A
21	119	6.3	1649	13	US-10-425-114-20370	Sequence 20370, A
22	115	6.1	3028	17	US-10-437-963-41007	Sequence 41007, A
23	115	6.1	3311	13	US-10-424-599-141176	Sequence 141176, A
24	114.8	6.1	1892	13	US-10-424-599-63874	Sequence 63874, A
25	114.4	6.1	1884	9	US-09-938-842A-905	Sequence 905, App
26	114.4	6.1	1884	11	US-09-938-842A-905	Sequence 905, App
27	114.4	6.1	1983	13	US-10-412-6998-637	Sequence 637, App
28	114.4	6.1	1983	13	US-10-225-066A-819	Sequence 819, App
29	114.4	6.1	1983	15	US-10-286-264-17	Sequence 17, Appl
30	114.4	6.1	1983	16	US-10-374-780A-2481	Sequence 2481, Ap
31	114.4	6.1	2008	13	US-10-424-599-21587	Sequence 21587, A
32	113.8	6.0	1637	13	US-10-425-114-13359	Sequence 13359, A
33	113.8	6.0	1645	13	US-10-424-599-65908	Sequence 65908, A
34	113.4	6.0	2545	13	US-10-412-6998-267	Sequence 267, App
35	113.4	6.0	2545	16	US-10-225-068-235	Sequence 235, App
36	113.4	6.0	2545	16	US-10-374-780A-2329	Sequence 2329, Ap
37	112.2	5.9	2517	17	US-10-437-963-3700	Sequence 3700, Ap
38	111.8	5.9	1179	13	US-10-425-114-7712	Sequence 7712, Ap
39	111.8	5.9	1546	13	US-10-425-114-29785	Sequence 29785, A
40	111.8	5.9	2736	13	US-10-424-599-37574	Sequence 37574, A
41	111.4	5.9	1237	17	US-10-437-963-50722	Sequence 50722, A
42	111	5.9	672	13	US-10-425-114-19291	Sequence 19291, A
43	111	5.9	1005	13	US-10-425-114-22835	Sequence 22835, A
44	111	5.9	1727	13	US-10-425-114-34958	Sequence 34958, A
45	111	5.9	2486	13	US-10-424-599-55280	Sequence 55280, A

ALIGNMENTS

RESULT 1

US-10-424-599-50272
; Sequence 50272, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION: Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 50272
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_16404C.1
US-10-424-599-50272

Query Match	9.1%	Score 171.6;	DB 13;	Length 422;
Best Local Similarity	72.5%	Pred. No. 8.2e-34;		
Matches 222;	Conservative 0;	Mismatches 84;	Indels 0;	Gaps 0;
Qy	1337	GAGCAAGACAAAGACTCAAGAACCTCTATGTTCCACACAGCATTCGCTTTTCAGCAGC	1396	
Db	25	GAGAAAGATAGAACCTTTGAAACATCGTTTATCAGAACAAATGGGCTTTCCACAGC	84	
Qy	1397	TGAACCAAGAGACCATCAGATTGGAGACCTCAAGAGGTTTGGCTTGAGAAATCTGTTT	1456	
Db	85	TGAGAGAGACGACGACAAATCTTGGAGGCTCAGAGGGGCTCGCTGAGAAATCTGTCT	144	
Qy	1457	CGGTTCTACGGAATTGATGTTTCCAAACCTTCTTACCTTACCCGAAAGATTCGGAGA	1516	

Db 145 CGGTTCTACGTGATGTTCCAAACTTCTTCCACCGTACCCAAAGACAACGAAA 204
Qy 1517 AACRTCTTCTAGCTATACGAGTGGCTTGACAAGAGTCAGGTATCAAACTGGTTTATAA 1576
Db 205 AGCATTTGCTAGCCATTCAAGTGGATTGACGAGAGCCAGGTGTCACAACTGGTTTATAA 264
Qy 1577 ATGCGCGGTTAGGCTATGGAAGCCGATGATAGAGAGATGATCGGAAATGAACAAGA 1636
Db 265 ATGCGAGAGTTTCGCTCTGGAACCATTTGATAGAGGAATGCTCGGAGGTAAACAAA 324
Qy 1637 GGAAGC 1642
Db 325 AGAATC 330

RESULT 2

US-10-425-114-13353
; Sequence 13353, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 13353
; LENGTH: 1555
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: JC-GMST02400042B10_FLI
US-10-425-114-13353

Query Match 6.9%; Score 130.2; DB 13; Length 1555;
Best Local Similarity 67.5%; Pred. No. 1e-22;
Matches 183; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
Qy 1370 TCACACAGCATTCGCTTCTTCCAGCAGCTCAACGAAAGCAACATCAGATTTGGAGCCTC 1429
Db 411 TCAGGCAACAAGGCGCACTTCAGCAGCTAGGAATGATTCACCCCAATGCATGGAGGCCCC 470
Qy 1430 AACGAGGTTTGCTGAGAAATCTGTTTCGGTTCTACGGAATGGATGTTCCAAAACCTCC 1489
Db 471 AAAGAGGCTTGCTGACGAGCTGTTTCAATCTTCGGGCTTGCTTTTGAGCATTTCC 530
Qy 1490 TTCACCTTTACCGAAAGATTCGGAGAAACATCTTCTAGCTATACGAGTGGCTTGACAA 1549
Db 531 TTCACCTTTATCCAAAGGACTCCGATAAAGTATGCTTGTAAACAAACTGGACTTGCTA 590
Qy 1550 GAAGTCAGGTATCAAACTGGTTTAAATGCGCGGTTAGGCTATGGAAGCCGATGATAG 1609
Db 591 GGAGCCAGGTGCAAACTGGTTTATCAATGCCGAGTTCGCTTTGGAGGCAATGGTTG 650
Qy 1610 AAGAGATGATCGGAAATGAACAAGAGAA 1640
Db 651 AAGAAATGTACTTGAAGAATAATCAAGGAGCA 681

RESULT 3

US-10-437-963-7710/c
; Sequence 7710, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 7710
; LENGTH: 1995
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_1427C.1
US-10-437-963-7710

Query Match 6.9%; Score 130.2; DB 17; Length 1995;
Best Local Similarity 65.1%; Pred. No. 1.2e-22;
Matches 192; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
Qy 1342 AAGACAAGACTCAAGAAACCTCTATGTTCCACCAGCATTCGCTTCTTCAGCAGCTGAAA 1401
Db 510 AAGAGCGGAGCTGGAGTGTGCTTCATCCAGAGCACTGGCGATGCGAGCTCCGG 451
Qy 1402 CGAAGAACCATCAGATTTGGAGACTCAACAGAGTTTCCTCGAGAAATCTCTTTGGTT 1461
Db 450 CGAGGTGACAGCAGTCGTGGCGCGCGCAGCGCGCTTCCCGAGAAAGTCCCGCGTG 391
Qy 1462 CTACGGAATGGATGTTCCAAAACCTCTTCAACCTTACCCCGAAGATTCGAGAAACAT 1521
Db 390 CTCAGGCTGGATGTTTCAGAACTTCTCCCGCGATTCCAAAGACAGCGAGAGGAC 331
Qy 1522 CTTCTAGCTATACGAGTGGCTTGACAAGAGTCAGGTATCAAACTGGTTTATAAATCG 1581
Db 330 ATGCTGCGCGGAGAGCGCTGAGCAGGAGCCAGGTCTCAACTGGTTTCAAAACGCC 271
Qy 1582 CGGTTAGGCTATGGAAGCCGATGATAGAGAGATGATGCGGAATGACAAAGA 1636
Db 270 CGCGTCCGACTGTGAAGCCGATGATCGAGGACATGTACGAGGAGCTCAAGAAGA 216

RESULT 4

US-10-424-599-130628
; Sequence 130628, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 130628
; LENGTH: 3033
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_88966C.1
US-10-424-599-130628

Query Match 6.9%; Score 130.2; DB 13; Length 3033;
Best Local Similarity 67.5%; Pred. No. 1.6e-22;
Matches 183; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
Qy 1370 TCCACCAGCATTCGCTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTTTCGAGACCTC 1429

Db 1374 TCAGGCAACAAGGCACTTCACGACTAGGATGATTCACCCCAATGCATGGAGGCC 1433
QY 1430 AAGAGGTTGGCTGAGAAATCTGTTTCGGTCTACGGAATGGATGTTCCAAAACCTCC 1489
Db 1434 AAGAGGCTTGCTGAAAGAGCTGTTCCATCTTCGGGCTTGGCTTTTGAGCATTC 1493
QY 1490 TTCACCTTACCGGAAGATTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAA 1549
Db 1494 TTCACCTTATCCAAAGGACTCGATAAAGTTATGCTTCTAAACAACCTGACTTGTA 1553
QY 1550 GAAGTCAGTATCAAACTGGTTTAAATGCGCGGTTAGGCTATGGAAGCCGATGATAG 1609
Db 1554 GGAGCCAGGTGTCAAACTGGTTTATCAATGCCGAGTTCGGCTTTGGAAGCCAAATGGTTG 1613
QY 1610 AAGAGATGATCGGGAATGACACAGAGAA 1640
Db 1614 AAGAAATGACTGGGAAGAAATCAAGGAGCA 1644

RESULT 5
US-10-437-963-31585
; Sequence 31585, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 14220
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(592)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB9829-031-Q6-K6-E4
US-10-021-323-14220

Query Match 6.7%; Score 126.4; DB 17; Length 2799;
Best Local Similarity 68.4%; Pred. No. 1.5e-21;
Matches 175; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1375 CAGCATTCGCTTCACGAGCTGAAACGAAAGACCAATCAGATTTGGAGACCTCAACGA 1434
Db 1590 CAACAGCGGCTTCCAGCAGTATGTTTGTACAGCAAAATGCTTGGAGGCCACAGAGG 1649
QY 1435 GGTTCCTCGTGAATACTGTTTCGGTCTACGGAATGGATGTTCCAAAACCTTCCTTCAC 1494
Db 1650 GACTGCCGGAACCTCAGTTCAATCTTCGTGCTTGGCTTTTGAACACTTCCTTCAC 1709
QY 1495 CTTACCCGAAAGATTCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAAGT 1554
Db 1710 CCGATCCAAAGATTCAGAAAGCTGATGCTAGCAGACAACTGGCTTAAACAAGT 1769
QY 1555 CAGGTATCAACTGGTTTAAATGCGCGGTTAGGCTATGGAAGCCGATGATAGAGAG 1614
Db 1770 CAGATTCAATGGTTTATAAATGCCGTTATGCCGTGTCGGCTGTGAAACCGATGATCGAAGAC 1829
QY 1615 ATGTATCGGAAATGA 1630
Db 1830 ATGTATAAAGAGAGA 1845

RESULT 6
US-10-021-323-14220

; Sequence 14220, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 14220
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(592)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB9829-031-Q6-K6-E4
US-10-021-323-14220

Query Match 6.7%; Score 126.2; DB 17; Length 592;
Best Local Similarity 68.7%; Pred. No. 5e-22;
Matches 173; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 1375 CAGCATTCGCTTCACGAGCTGAAACGAAAGACCAATCAGATTTGGAGACCTCAACGA 1434
Db 27 CAACANAGAGCTCTTCAGCAGTTTGGTATGATGATCAACAACATGATGAGGCGCTCAGAGA 86
QY 1435 GGTTCCTCGTGAATACTGTTTCGGTCTACGGAATGGATGTTCCAAAACCTTCCTTCAC 1494
Db 87 GGGCTGCCGAAGCTCTGTTTCAATCTTCGTCTGGCTATTTGAGCATTTTCTTCAT 146
QY 1495 CTTACCCGAAAGATTCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAAGT 1554
Db 147 CCCTACCCGAAGGATTCGTATAGATCATCTAGCAAGACAGACAGGCTTAACTAGAAGT 206
QY 1555 CAGGTATCAACTGGTTTATAAATGCCGCTTGGCTATGGAAGCCGATGATAGAGAG 1614
Db 207 CAGTTTCGAATGGTTTATAAATGCTCGGGTGGCTCTTTGGAAGCCCATGGTGAAGAG 266
QY 1615 ATGTATCGGAA 1626
Db 267 ATGTACAGGA 278

RESULT 7
US-10-425-114-9882
; Sequence 9882, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 9882
; LENGTH: 1716
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:

OTHER INFORMATION: Clone ID: 700875318_FLI
US-10-425-114-9882

Query Match 6.6%; Score 125.2; DB 13; Length 1716;
Best Local Similarity 66.9%; Pred. No. 2.2e-21;
Matches 178; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 1370 TCACACAGATTGCCCTTCTTACAGCAGCTGAACGAAAGAACCATCAGATTTGGAGACCTC 1429
DB 213 TTCCGCAACAAGGCGCACTTCAGCAACTAGGAATGATTCAGCCCAATCATGGAGGCCCC 272
QY 1430 AACGAGTTTGCTCAGAAATCTGTTGGTTCTACGGAATTTGGATGTTCCAAAATCTCC 1489
DB 273 AGAGAGCTTGCTGACAGAGCTGTTTCATTTCTCGGGTGGCTTTTGGAGCATTTCC 332
QY 1490 TTCACCCCTTACCAGAAAGATTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAA 1549
DB 333 TTCACCCATATCCAAAGGACTCCGATAAAGTTATGCTTAAACAAACTGGACTTTCTA 392
QY 1550 GAAGTCAGCTATCAACTGTTTATAAATGCGCGGTTAGGCTATGGAAGCCGATGATG 1609
DB 393 GGAGCAGGTGCAAACTGGTTTATCAATGCCGAGTTCGGCTTTGGAAGCCAAATGGTTG 452
QY 1610 AAGAGATGTATCGGAAATGAACAAG 1635
DB 453 AAGAAATGTACTTGAAGAATCAAG 478

RESULT 8

US-10-425-114-5047
; Sequence 5047, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 5047
; LENGTH: 1452
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700441126_FLI
US-10-425-114-5047

Query Match 6.6%; Score 124.4; DB 13; Length 1452;
Best Local Similarity 72.5%; Pred. No. 3.1e-21;
Matches 161; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 1409 ACCATCAGATTGGAGACCTCAACGAGGTTTCCTCGAGAAATCTGTTTCGGTTCTACGGA 1468
DB 316 AGCAAAACCTTGGAGGCCACAGAGGACTACCTGAAATTCAGTCTCGATTCTCCGGG 375
QY 1469 ATTGATGTTCCAAACTCTTCCCTTACCCCTTACCCGAAAGATTCGGAGAAATCTTCTAG 1528
DB 376 CTGCGCTTTTGAACACTTCTTCCCTTACCCCTGATCCAAAGATTTGAAAGCTAATGTTAT 435
QY 1529 CTATACGAAGTGGCTTGACAAGAAGTCAGGTATCAAACTGGTTTATAAATCGCGGGTTA 1588
DB 436 CGAGACAAACTGGGTTGACAAGAAGTCAGATTCAAACTGGTTTATCAATGCCGCTGCC 495
QY 1589 GGCTATGGAGCCGATGATAGAGAGATGATGCGGAAATGA 1630
DB 496 GGCTGTGGAAGCCAAATGATCGAAGACATGTACAAAGAAGAGA 537

RESULT 9

US-10-425-114-7804
; Sequence 7804, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 7804
; LENGTH: 1606
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700684286_FLI
US-10-425-114-7804

Query Match 6.6%; Score 124.4; DB 13; Length 1606;
Best Local Similarity 65.5%; Pred. No. 3.4e-21;
Matches 182; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 1375 CACCATTCGCTTCTTCCAGCAGTGAACGAAAGAACCATCAGATTTGGAGACCTCAACGA 1434
DB 381 CAACAACGTCATTAACAACAGCTTGGATGATCAACAACATCTTGGAGGCCCAAGA 440
QY 1435 GGTTCGCTCAGAAATCTGTTTCGGTTCTACGGAATTTGGATGTTCCAAAATCTTCTTAC 1494
DB 441 GGCTTCGCTCAGCTGCTGTTTCTGTTCTTTCGAGCTTGGCTTTTGAACATTTCTTGAC 500
QY 1495 CTTTACCCGAAAGATTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGAAGAAGT 1554
DB 501 CCTTATCCTTAAGACTCCGATTAAGGTTATGCTTGCAAAACAACTGGACTTACTCGGAGC 560
QY 1555 CAGGTATCAACTGGTTTATAATGCGCGGTTAGGCTATGGAAGCCGATGATAGAGAG 1614
DB 561 CAGGTGCTCAACTGGTTTATAATGCGCGGTTAGGCTATGGAAGCCCAATGTTAGAGAA 620
QY 1615 ATGTATGCGGAAATGAACAAGAGGAACTCAATAACAG 1652
DB 621 ATGTACTTGGAGAGGTAAACAAGAACCAACATAG 658

RESULT 10

US-10-437-963-19673/c
; Sequence 19673, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 19673
; LENGTH: 1785
; TYPE: DNA

; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_2510C.1
US-10-437-963-19673

Query Match 6.6%; Score 124.2; DB 17; Length 1785;
Best Local Similarity 65.1%; Pred. No. 4.1e-21;
Matches 183; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1357 GAAACCTCTATGTTCCACAGACATTCCTCTTTCAGCAGCTGAAACGAAAGAACCATCG 1416
DB 399 GAGTCGGCGTTATCCAGAAAGCACTGGCGGTGCAGCAGCTCGGCGCGGAGCAGCAG 340
QY 1417 ATTGTGAGACCTCAACGAGGTTTGCTGAGAAATCTGTTTCGTTCTACGGAATGGATG 1476
DB 339 TGTGCGCGCGCAGCGCGCTCCGAGAGAGTCGTCGCGTCTCAAGSCCTGGATG 280
QY 1477 TTCCAAAATCTCTTCCACCTTACCCGAAAGATTGCGAAGAACATCTTCTAGCTATACGA 1536
DB 279 TTGAGAATCTCTCCGCGCGTATCCGAAAGCAGCGAGAGGATGCTCGCAGCAAGG 220
QY 1537 AGTGGCTTGACAGAGAGTCAGGTATCAAACTGGTTTATAAATGCGCGGTTAGGCTATGG 1596
DB 219 AGCGGCTCAGAGGAAACAGGTGTCGAATGTTTCATAAAGCTCGAGTTCGGTATGG 160
QY 1597 AAGCCGATGATGAAGAGATGTATGCGGAAATCAACAAGAG 1637
DB 159 AAGCCAATGATCAGGAGATGTCGAGGAGCTGAAGAGGAG 119

RESULT 11
US-09-938-842A-1337
; Sequence 1337, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1337
; LENGTH: 2043
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1337

Query Match 6.6%; Score 123.6; DB 9; Length 2043;
Best Local Similarity 69.4%; Pred. No. 6.3e-21;
Matches 168; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1410 CCATCAGATTGGAGACCTCAACGAGGTTTGCTGAGAAATCTGTTTCGGTTCTACGGAA 1469
DB 1146 CAATAATGCTTGGAGACCTCAACGCTGCTCCAGACAGCGCTCTCAGTTCTCGGTGC 1205
QY 1470 TTGGATGTTCCAAAATCTCTTACCCCTTACCCGAAAGATTGCGAAGAACATCTTCTAGC 1529
DB 1206 TTGGCTCTTCGAACACTTTCTTCATCCATACCCCTAAGGATTCGGACAAGCATGCTAGC 1265
QY 1530 TATACGAAGTGGCTTCACAGAGATCAGGTATCAAACTGGTTTATAAATGCGCGGTTAG 1589
DB 1266 TAAGCAACAGGACTCACTCGTAGCCAGGTGTCGAACCTGGTTTATAAACCAGGAGATTGC 1325

QY 1590 GCTATGGAAGCCGATGATAGAGAGATGTGCGGAAATGAACAAGAGAGCTCAATAA 1649
DB 1326 GTTATGGAACCAATGTTGGAGGAGATGTACATGGAGAAATGAAGGAGCGCAAGAA 1385
QY 1650 CA 1651
DB 1386 CA 1387
RESULT 12
US-09-938-842A-1337
; Sequence 1337, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1337
; LENGTH: 2043
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1337

Query Match 6.6%; Score 123.6; DB 11; Length 2043;
Best Local Similarity 69.4%; Pred. No. 6.3e-21;
Matches 168; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 1410 CCATCAGATTGGAGACCTCAACGAGGTTTGCTGAGAAATCTGTTTCGGTTCTACGGAA 1469
DB 1146 CAATAATGCTTGGAGACCTCAACGCTGCTCCAGACAGCGCTCTCAGTTCTCGGTGC 1205
QY 1470 TTGGATGTTCCAAAATCTCTTACCCCTTACCCGAAAGATTGCGAAGAACATCTTCTAGC 1529
DB 1206 TTGGCTCTTCGAACACTTTCTTCATCCATACCCCTAAGGATTCGGACAAGCATGCTAGC 1265
QY 1530 TATACGAAGTGGCTTCACAGAGATCAGGTATCAAACTGGTTTATAAATGCGCGGTTAG 1589
DB 1266 TAAGCAACAGGACTCACTCGTAGCCAGGTGTCGAACCTGGTTTATAAACCAGGAGATTGC 1325
QY 1590 GCTATGGAAGCCGATGATAGAGAGATGTGCGGAAATGAACAAGAGAGCTCAATAA 1649
DB 1326 GTTATGGAACCAATGTTGGAGGAGATGTACATGGAGAAATGAAGGAGCGCAAGAA 1385
QY 1650 CA 1651
DB 1386 CA 1387

RESULT 13
US-10-225-066A-335
; Sequence 335, Application US/10225066A
; Publication No. US20030226173A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: RIECHMANN, Jose Luis
; APPLICANT: ADAM, Luc J
; APPLICANT: DUBELL, Arnold T
; APPLICANT: HEARD, Jacqueline E
; APPLICANT: PILGRIM, Marsha L

[illegible]

```

; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: US 09/229,413
; PRIOR FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 3853
; SEQ ID NO 2000
; LENGTH: 227
; TYPE: DNA
; ORGANISM: Glycine max
US-09-922-293-2000

      6.5%;      Score 122.6;  DB 12;      Length 227;
      Best Local Similarity 72.8%;  Pred. No. 2.8e-21;
      Matches 158;  Conservative 0;  Mismatches 59;  Indels 0;  Gaps 0;

1408  AACATACGATTTGGAGACCTCAACGAGGTTTGCTTGAGAAATCTCTTCGGTTCTACGG 1467
      5  AATCACAATGCTGGAGACCCACAGAGAGGATTGCCTGAGAGATCGGTTTCTGTCTCGCGT 64

1468  AATTGGATGTTCCAAACATTCCTTACCCCTTACCCGAAAGATTCGAGAAACATCTCTA 1527
      65  GCTTGSGCTCTTTGAACACATTTCTCCACCCCTTATCCCAAGGACTCGGACAAACACATGCTT 124

1528  GGTATACGAAGTGGCTTTGACAAAGATCGAGTATCAAACTGGTTTATAATGCGGGGTT 1587
      125  GCACAAACAAACAGGGCTCACCAGAAGCCAGGTTTCAANTTGGTTTCATAATGCTCGAGTT 184

1588  AGGCTATGGAAGCCGATGATAGAAGAGATGTATGCGG 1624

185  CGGCTTTTGAAGCAATGGTGGCGGCATGTACTTGG 221

```

Search completed: September 2, 2004, 12:02:33
Job time : 889 secs

Db 21 AsnProLeuMetAspPheileProSerArgGluAspSerThrSerPheSerThrMet 40
 QY 424 CTTCCATGGAATACCATCAGATCAGATCTCTCAAAATGGGTGGCTTTGATATTTCAAT 483
 Db 41 LeuProTrpAsnThrIleArgSerAspProLeuGlnMetGlyGlyPheAspIlePheAsn 60
 QY 484 TCTATGCTGACTAAACAAATCTATCTCTCCACGCTCTATCGATGTTCAAGATAAC 543
 Db 61 SerMetLeuThrAsnLysThrLeuSerSerProArgSerIleAspValGlnAspAsn 80
 QY 544 CGCAATGTTGAGTTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 603
 Db 81 ArgAsnValGluPheMetAlaProProHisProProLeuHisProLeuAspHis 100
 QY 604 TTAAGACACTATGATGATCTCTCAACACATGCTGGGTGTTTGAAGCAATAGTGGTTT 663
 Db 101 LeuArgHisThrAspAspSerSerAsnAsnMetIlePheGluAlaAsnSerGluPhe 120
 QY 664 CAGGCAATTTTCAGGTGATGTTGGTCCAAAGTGAACCAATGATGCTCTACATTCGGTGAAGAA 723
 Db 121 GlnAlaPheSerGlyValValGlyProSerGluProMetSerThrPheGlyGluGlu 140
 QY 724 GATTTCCCGTCTCTAATTTGGAATAAAGAAACATGAGCTTCTCTAGTCTTGATCA 783
 Db 141 AspPheProPheLeuIleSerAsnLysArgAsnAsnGluLeuSerLeuSerLeuAlaSer 160
 QY 784 GATGTTCTGATCAATGCTCGAGATAAGTCTTCTGCGAGCTTACAAAGATTAGCCTCAGAG 843
 Db 161 AspValSerAspGluCysSerGluIleSerLeuCysAlaAlaThrArgLeuAlaSerGlu 180
 QY 844 CAAAGTCTTTCAGCAGCAAGAACATTTCTAATAACGTTGTTACTCAAGGTTTCTCTCAA 903
 Db 181 GlnAlaSerCysSerSerLysAspIleSerAsnAsnValThrGlnGlyPheSerGln 200
 QY 904 CTTATATTGGCTCAAAATACCTTCACTCTGTTCAAGAAATACTATCTCATTTCGCCGCA 963
 Db 201 LeuIlePheGlySerLysThrLeuHisSerValGlnGluIleLeuSerHisPheAlaAla 220
 QY 964 TACTCGCTCGATTATTCATCTCAGAGAACCGAGTCCAGGAGTGTGATTCAGCCTTACT 1023
 Db 221 TyrSerLeuAspTyrSerSerArgGlyThrGluSerGlyAlaAlaSerSerAlaPheThr 240
 QY 1024 TCACGTTTTCAGAAATACATGATGTTCTGATGTTGATGTTCTAATACTCCGAGCGGTT 1083
 Db 241 SerArgPheGluAsnIleThrGluPheLeuAspGlyAspSerAsnAsnSerGluAlaGly 260
 QY 1084 TTCGGATCTTACATTTCAAAGGAGAGCATTTAGAAGCAAGAAACCCATCTCTCTGGATCTT 1143
 Db 261 PheGlySerThrPheGlnArgAlaLeuGluAlaLysLysThrHisLeuLeuAspLeu 280
 QY 1144 CTTCAAAATGGTGGATGATGATAGTATGATTCATTCGCTAGATGATCAGCTTATATCA 1203
 Db 281 LeuGlnMetValAspAspArgTyrSerHisCysValAspGluIleHisThrValIleSer 300
 QY 1204 GGTTCATGCTGCAACCGAGTTAGATCCACAGTTTACACCCCGTTTGGCTCCCAAC 1263
 Db 301 AlaPheHisAlaAlaThrGluLeuAspProGlnLeuHisThrArgPheAlaLeuGlnThr 320
 QY 1264 GTTTCCTCTTTTACAAGAACCTGAGAGAGAGAAATCTCTCAAGAGATTAATCTTATGGGA 1323
 Db 321 ValSerPheLeuTyrLysAsnLeuArgGluArgIleCysLysLysIleSerMetGly 340
 QY 1324 TCTGTATTGGAGAGGCAACACAGACTCAAGAAACCTCTATGTTCCACGAGATTGC 1383
 Db 341 SerValLeuGluArgGlyLysAspLysThrGlnGluThrSerMetPheHisGlnHisCys 360
 QY 1384 CTTCTTCAGCACTGAAACGAAAGAACCATCATGATTTTCGAGACCTCAACGAGGTTGCTT 1443
 Db 361 LeuLeuGlnGlnLeuLysArgLysAsnHisGlnIleTrpArgProGlnArgGlyLeuPro 380
 QY 1444 GAGAAATCTGTTTCGGTCTACGGATGTTGATGTTCCAAACTCTTCACCCCTTACCCG 1503
 Db 381 GluLysSerValSerValLeuArgAsnTrpMetPheGlnAsnPheLeuHisProTyrPro 400

QY 1504 AAGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGCACAGAGTCCAGGTATCA 1563
 Db 401 LysAspSerGluLysHisLeuLeuAlaIleArgSerGlyLeuThrArgSerGlnValSer 420
 QY 1564 AACTGCTTTTATAAATCCGGGGTTAGCTATGAGCCCATGATGAGAGATGATGCG 1623
 Db 421 AsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluGluMetTyrAla 440
 QY 1624 GAAATGAACAGAGGAGCTCAATAACAGTCAATCAACCAACCGACCACTCTTCCA 1683
 Db 441 GluMetAsnLysArgLysLeuAsnAsnSerHisIleGlnProAsnGlyProThrLeuArg 460
 QY 1684 ATCCCAAAATCTGTTATGATGAGCCAAAGCAATGCATAAA 1722
 Db 461 MetProLysSerValMetMetSerGlnAlaMetHisLys 473
 RESULT 2
 H84774
 probable homeodomain transcription factor [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: H84774
 R.;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; A.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.E.
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: H84774
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-680 <STO>
 A:Cross-references: GB:AE002093; NID:g4510375; PIDN:AAD21463.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g35940
 A:Map position: 2
 Alignment Scores:
 Pred. No.: 2,65e-29 Length: 680
 Score: 472.00 Matches: 152
 Percent Similarity: 42.01% Conservative: 74
 Best Local Similarity: 28.25% Mismatches: 193
 Query Match: 14.37% Indels: 119
 DB: 13 Gaps: 13
 US-09-423-575-1 (1-1886) x H84774 (1-680)
 QY 283 ATAAAGTTTCATAGAAACCAATGGACAAACAAACAAACACACTTTTAGTTCTCTG 342
 Db 32 ValGlnTyrThrGlnGlnAspAsnAspSerAsnAsnAsnAsnAsnSerAsnAsnSerAsn 51
 QY 343 GAT-----AATGTCATGACTAACCAAAATCCCTCTCTCATGATTTTATACCTCA 393
 Db 52 AsnAsnAsnThrAsnThrAsnThrAsnAsnAsnSerSerPheValPheLeuAspSer 71
 QY 394 AGAGAAGATTCACCTTCATTCACCAATGCTTCCATGGAATACCATCAGATCAGATCCT 453
 Db 72 HisAlaPro-GlnProAsnAlaSerGlnPheValGlyIleProLeu----- 87
 QY 454 CTACAAATGGTGGCTTTCATATTTTCAATTCTATGCTGACTAACAAATACCTATCATCT 513
 Db 88 -----SerGlyHis-- 90
 QY 514 TCTCCAGCTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTCATGCTCTCTCTCTCT 573
 Db 91 GluAlaAlaSerIleThrAlaAlaAspAsnIleSerValLeuHisGlyTyrProProArg 110
 QY 574 CATCCCTCCCTCACTTCATCTCTTGGATCATTTTAAGACACTATGATGATTCCTCAACAC 633
 Db 111 ValGlnTyrSerLeuTyrGly-----SerHisGlnValAspProThrHisGln 126
 QY 634 ATGTGGGGTTTCAACCAATAGTAGTTTCAGGCATTTTCAGGTGTAGTTGGTCCAGT 693


```

QY 687 TCCAGTGAACCAATGATGCTCTACATTCGGTGAAGAAGATTTCCCGTTTCTTAATTGCA 746
Db 117 -----SerThrMetSerAspGlu-----
QY 747 TAAAGAAACATGAGCTTTTCATTTGAGCTTCGATCAGATGTTTCTGATGAATGCTCGGA 806
Db 123 -----AsnGlySerLeuSerValHisGlnHisSerAspGlnIleLeuPr 139
QY 807 GATAGTCTTTGTGCAGCTACAAGATTAGCTCAGAGCAAGCTTCTTCAGCAGCAAGA 866
Db 139 oSerSerVal-----TyrAsnAs 145
QY 867 CATTTCTAATAACGTTGTT-----ACTCAAGTTTCTC 899
Db 145 nAsnGlyAsnAsnGlyValGlyPheTyrAsnAsnTyrArgTyrGluThrSerGlyPheVa 165
QY 900 TCAACTTATATTGGCTCAAAATACCTTCACCTGTTCAAGAAATACATCTCATTTTCGC 959
Db 165 lSerSerValLeuArgSerA-gTyrLeuLysProThrGlnGlnLeuLeuAspGluValVa 185
QY 960 CGCATCTCGCTCGATTAT-----TCATCTCGAGGAACCGA 995
Db 185 lSerValArgLysAspLeuLysLeuGlyAsnLysLysMetLysAsnAspLysGlyGlnAs 205
QY 996 GTCAGGAGCTGCTAGTTTCAGCTTACTTCACGTTTTCAGAAATATACTAGTTTCTTGA 1055
Db 205 pPheHisAsnGlySerSer-----AspAsnIleThrGlu-----As 217
QY 1056 TGGTCAATCTAATAATCGGAGCGGGTTCGGATCTACATTTCAAAGGAGGACATTAAGA 1115
Db 217 pAspLysSerGlnSerGlnGluLeu-----SerProSerGluArgGlnGluLeuGl 234
QY 1116 AGCAAGAAACCCATCTCTTGGATCTTCTCAATGTTGATGATGATGATGATGATGATG 1175
Db 234 nSerLysLysSerLysLeuLeuThrMetValAspGluValAspLysArgTyrAsnGinty 254
QY 1176 CGTAGATGAGATTCATACGGTTATATACAGCTTCATCTCCACCGAGTTAGAT---CC 1232
Db 254 rHisGlnMetGluAlaLeuAlaSerPheGluMetValThrGlyLeuGlyAlaAl 274
QY 1233 ACAGTTACACACCGGGTTGGCTCCCAACCGCTTCTTCTTACAGAACCTGAGAGA 1292
Db 274 aLysProTyrThrSerValAlaLeuAsnArgIleSerArgHisPheArgCysLeuArgAs 294
QY 1293 GAGATCTGCAAGAGATA---ATCTCTATGGATCTGTATTGAGAGAGCGCAAGACAA 1349
Db 294 pAlaIleLysGluGlnIleGlnValIleArgGlyLysLeuGlyGluArgGluThrSerAs 314
QY 1350 GACTCAAGAAACCTCTATGTTCCACCAAGCAATGCTTCTTCAGCAGCTGAACAGAA 1409
Db 314 pGluGlnGlyLysArgIleProArgLeuArgTyrLeuAspGlnArgLeuArgGlnGlnAr 334
QY 1410 C-----CATCAG-----ATTGGAGACTCAACGAGGTTTGC 1442
Db 334 gAlaLeuHisGlnGlnLeuGlyMetValArgProAlaTrpArgProGlnArgGlyLeuPr 354
QY 1443 TGAGAAATCTGTTTCGGTTCTACGGAATTTGGATGTTCCAAATCTCTTCCACCTTACCC 1502
Db 354 oGluAsnSerValSerIleLeuArgAlaTrpLeuPheGluHisPheLeuHisProTyrPr 374
QY 1503 GAAAGATTCGAGAAACATCTCTAGCTATACGAAGTGGCTTGACAGAGTCAAGTATC 1562
Db 374 oLysGluSerGluLysIleMetLeuSerLysGlnThrGlyLeuSerLysAsnGlnValAl 394
QY 1563 AAACCTGGTTTAAATCGCGGTTAGCTATGAAGCCGATGATAGAAGATGTATGC 1622
Db 394 aAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluGluMetTyrLy 414
QY 1623 GGAA-----ATGACAAGAGGAAGCTCAATAACAGTCAATTCACCCAAACGGACC 1673
Db 414 sGluGluPheGlyGluSerAlaGluLeuLeuSerAsnSer-----AsnGlnAs 430
```

```

QY 1674 AACTCTTCGAATGCCAAATCTGTTATGATGAGCCAGCAATGCATATAA----- 1722
Db 430 pThrLysLysMetGlnGluThr-----SerGlnLeuLysHisGluAspSerSerSe 447
QY 1723 -----TAAGACAAACAATGTGTTTACCAACTTTGTGATAATTA 1760
Db 447 rSerGlnGlnGlnAsnGlnGlyAsnAsnAsnAsnIleProTyrThrSerAsp-AlaG 467
QY 1761 GGCATTTGCTACTCTATGATTCGCCAAACCTTAAACCATGTACGACT 1807
Db 467 luGlnAsnLeuValPheAlaAspProLysProAspArgAlaThrThr 482

RESULT 4
H86329
F69.25 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C:Accession: H86329
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719; PMID:11130712
A:Accession: H86329
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1538 <STO>
A:Cross-references: GB:AE005172; NID:g10086497; PIDN:AAG12557.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Alignment Scores:
Pred. No.: 1,22e-25 Length: 538
Score: 425.50 Matches: 119
Percent Similarity: 43.04% Conservative: 59
Best Local Similarity: 32.78% Mismatches: 314
Query Match: 12.96% Indels: 31
DB: 2 Gaps: 9

US-09-423-575-1 (1-1886) x H86329 (1-538)
QY 746 ATAAAGAAACATGAGCTTTTCATTTGATGATGATGATGATGATGATGATGATGATG 805
Db 111 ValLysGluThrSerProPheHis-----ValAspGluMet--SerValLysSerLysG 128
QY 806 AGATAAGTCTTTGTGCAGCTACAGATTTAGCCTCAGACCAAGCTTCTTCGACGACGAA 865
Db 128 luMetIleLeuLeuGlyGlnSerAspProSerSerGlyTyrAlaGlyAsnGlyLysG 148
QY 866 ACATTTCTTAATAACGTT-----GTTACTCAGGTTTCTCTCACTATATTG 913
Db 148 lYPheTyrAsnAsnTyrArgTyrAsnGluThrSerGlyGlyPheMetSerSerValLeuA 168
QY 914 GTCATAAATACCTTCACCTCTGTTCAAGAAATACATCTCTCCGCCGACTACTCGCTCG 973
Db 168 rSerArgTyrLeuLysProAlaGlnAsnLeuLeuAspGluValSerValLysLysG 188
QY 974 ATTATTATCTCGAGAACCGAGTCAGAGCTGCTAGT-----TCAGCCTTTACTTCA 1027
Db 188 luLeuAsnGlnMetGlyLysLysLysMetLysValAsnAspPheAsnSerGlySerLysG 208
QY 1028 GTTTTGAGATATACTGAGTTTCTTGATGGTGAATTAATAACTCGGAGCGGGTTTCG 1087
Db 208 luIleGluGlyGlyGlyGluLeuSerSerAspSerAsnGlyLysSerIleGluLeu- 227
QY 1088 GATCTACATTTCAAAGGAGAGCATTTAGAGCAAGAAACCATCTCTTGATCTCTTCTTC 1147
Db 1147 -----
```

```

Db      228 --SerThrIleGluArgGluGluLeuGlnAsnLysLysAsnLysLeuLeuThrMetValA 247
QY      1148 AATGCTGATGATCATATAGTATAGTCATGCTAGATGAGATTCATACGGTATATATACGGCT 1207
Db      247 spCluValAspLysArgGlyArgGlnTyrHisGlnMetGluAlaLeuAlaSerSerP 267
QY      1208 TCCATGCTCAACCGAGTTA--GATCCACAGTTACACACCGGTTTGCCCTCCAAACCG 1264
Db      267 heCluIleValAlaGlyLeuGlySerAlaLysProTyrThrSerValAlaLeuAsnArgI 287
QY      1265 TTTCCTTCTTATACAGAACCCTGAGAGAGAGATTCGCAAGAGATA---ATCTCTATGG 1321
Db      287 leSerArgHisPheArgAlaLeuArgAspAlaIleLysGluGlnIleGlnIleValArgG 307
QY      1322 GATCTGTATTGGAGAGAGCAAGACCAAGACTCAAGAAACCTCT----- 1365
Db      307 luLysLeuGlyGluLysGlyGlyGluSerLeuAspGluGlnGlnGlyGluArgIleProA 327
QY      1366 -----ATGTTCCACCAAGATTCGCTTCTTCCAGCAGCTGAAC 1402
Db      327 rgLeuArgTyrLeuAspGlnArgLeuArgGlnGlnArgAlaLeuHisGlnGlnLeuGlyW 347
QY      1403 GAAAGAACCATCAGATTTGGAGACCTCAACGAGTTTGCTGAGAAATCTGTTTCGGTTC 1462
Db      347 etVal---ArgProAlaIlePrgProGlnArgGlyLeuProGluAsnSerValSerValL 366
QY      1463 TACGGAATTTGGATGTTCCAAACTTCCCTTACCCCTTACCCGAAAGATTCGGAGAAACATC 1522
Db      366 euArgAlaIlePheGluHisPheLeuHisProTyrProLysGluSerGluLysIleW 386
QY      1523 TTCTAGCTATAGAGAGTGGCTTGACAGAGAGTCAAGATTCAGATTCAACTGTTTATATAGCGC 1582
Db      386 etLeuAlaLysGlnThrGlyLeuSerLysAsnGlnValAlaAsnTrpPheIleAsnAlaA 406
QY      1583 GGGTTAGGCTATGGAACCGATGATAGAGAGATGTATGCGGAA-----ATGA 1630
Db      406 rgValArgLeuTrpLysProMetIleGluGluMetTyrLysGluGluPheGlyAspGluS 426
QY      1631 ACAAGAGAGAGCTCAATACATCATTCATCAACCAAGCAACCACTCTTTCGAATGCCAA 1690
Db      426 exGluLeuLeuIleSerLysSerSerGlnGluProAsnSerThrAsnGlnGluAspSers 446
QY      1691 AATCTGTTATGATGACCAAGCAATGTCATAATAAGACAACTGTTGTACCAACTTT 1750
Db      446 erSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 466
QY      1751 GTGAT 1755
Db      466 laasp 467

RESULT 5
D85435
BEU1-like homeobox 2 protein (BLH2) [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: D85435
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: D85435
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-638 <STO>
A:Cross-references: GB:NC_001268; NID:97270636; PIDN:CAB80353.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4G36870
A:Map position: 4

Alignment Scores:
Pred. No.: 2,61e-25 Length: 638
Score: 421.50 Matches: 160
Percent Similarity: 41.12% Conservative: 90

```

```

Best Local Similarity: 26.32% Mismatches: 168
Query Match: 12.83% Indels: 190
DB: 2 Gaps: 26
US-09-423-575-1 (1-1886) x D85435 (1-638)
QY      289 TTTCATGAACCAACCAATGGACAACAAC----- 315
Db      17 PheHisAsnAsnSerMetSerGlnAspTyrHisHisHisHisAsnGlnHisGln 36
QY      316 -----AACAAACAACACATTTTAGTTCTCTGGAT-----AATGTCATGACT 357
Db      37 GlyGlyIlePheAsnPheSerAsnGlyPheAspArgSerAspSerProAsnLeuThrThr 56
QY      358 AACCAAAATCCT-----CTTCTCATCGAT----- 381
Db      57 GlnGlnLysGlnGluHisGlnArgValGluMetAspGluSerSerValAlaGlyGly 76
QY      382 TTATACATCTTCAAGAGAGAGATTCAACTTCACTTCTCAACAATGTTTCCATGCAATACCATC 441
Db      77 ArgIleProValTyrGluSerAlaGlyMetLeuSerGluMetPheAsnPhe----- 93
QY      442 AGATCAGATCTCTACAAATGGGTGGC-----TTTGATATTTTCAATTTCTATGCTG 492
Db      94 -----ProGlySerSerGlyGlyArgAspLeuAspLeuGlyGlnSerPheArg 110
QY      493 ACTAACCAATACTTATCATCTTCTCCACGGTCTATCATGTTCAAGATAAACCGCAATGTT 552
Db      111 SerAsnArgGlnLeuLeuGlu-----GluGlnHisGlnAsnIle 123
QY      553 GAGTTCATG----- 561
Db      124 ProAlaMetAsnAlaThrAspSerAlaThrAlaAlaAlaMetGlnLeuPheLeu 143
QY      562 -----GCTCCTCTCTCTCATCTCTCTCTCA----- 585
Db      144 MetAsnProProProGlnGlnProProSerProSerSerThrThrSerProArgSer 163
QY      586 -----CTTCAT-----CCTTGGATCATTTTAAGACAC 612
Db      164 HisHisAsnSerSerThrLeuHisMetLeuLeuProSerProSerThrAsnThrThrHis 183
QY      613 TATCATGATCTCTCAACAACATG----- 636
Db      184 HisGlnAsnTyrThrAsnHisMetSerMetHisGlnLeuProHisGlnHisGlnGln 203
QY      637 -----TGGGGTTTGAAGCAAAATAGTAGCTTTTCAGGCATTTTCAGGTAGTTGGT 687
Db      204 IleSerThrTrpGlnSerSerProAspHisHisHisHisHisHisAsn----- 219
QY      688 CCAAGTGAACCAATGATGCTCATCGGTGAAGAGATTTCCCGTTTCTAATTTCAAT 747
Db      220 ---SerGlnThrGluIleGlyThrValHisValGluAsn-----SerGlyGly 234
QY      748 AAAAGAAACAATGAGCTTTTCATTCAGTCTTTCATGATGATGATGATGATGATGATGATGAT 792
Db      235 HisGlyGlyGlnGlyLeuSerLeuSerLeuSerLeuSerLeuSerLeuSerLeuSerLeuSer 254
QY      793 GATGAATGCTCGGAGATAAGTCTTTGTGACAGTACAAAGATTAGCTCA----- 840
Db      255 GluGluTyrArgAsnIleTyrTyrGlyAlaAsnSerSerAsnAlaSerProHisHisGln 274
QY      841 -----GAGCAAGCTTTTCAGCAGCAGCAAGACATTTCTAATACGTTGTACT 888
Db      275 TyrAsnGlnPheLysThrLeuLeuAlaAsnSerSerGlnHisHisGlnHisGlnHisGln 294
QY      889 CAAGGTTTCTCT-----CAACTATATATTT 912
Db      295 GlnPheArgSerSerProAlaAlaSerSerSerMetAlaAlaValAsnIleLeuArg 314
QY      913 GGCTCAAAATACCTTCACTCTGTTCAAGAAATACTATCTCATTTCCCGCATCTCGCTC 972
Db      315 AsnSerArgTyrThrThrAlaAlaGlnGluLeuGluGluPheCysSer--ValGlyA 334

```


Db 272 GlnArgAlaLeuGlnArgLeuGlyValMetGlnProHisThrTrpArgProGlnArgGly 291
Qy 1438 TTGCTGAGAAATCTGTTTCGGTTCTACGGAATGGATGTTCACAAACTCTCTCCACCC 1497
Db 292 LeuProAspSerValLeuValLeuArgAlaTrpLeuPheGlnHisPheLeuHisPro 311
Qy 1498 TACCCGAAAGATTCCGAGAAACATCTCTAGCTATACGAAGTGGCTGCACAAAGATCAG 1557
Db 312 TyPProLysAspSerAspLysIleMetLeuAlaArgGlnThrGlyLeuSerGlyGln 331
Qy 1558 GTATCAAACTGGTTTATAAATGCGCGGTTAGCTATGGAAGCCGATGATAGAGAGATG 1617
Db 332 ValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetValGluGluMet 351
Qy 1618 TATCGCGAA 1626
Db 352 TyTrpLysGlu 354
RESULT 7
C84670
Probable homeodomain transcription factor [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84670
R.; Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; PMID:20083487; PMID:10617197
A:Accession: C84670
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-431 <STO>
A:Cross-references: GB:AE002093; NID:G6598540; PIDN:AAF18625.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g27220
A:Map position: 2
Alignment Scores:
Pred. No.: 5,04e-25 Length: 431
Score: 417.50 Matches: 102
Percent Similarity: 50.83% Conservative: 51
Best Local Similarity: 33.89% Mismatches: 91
Query Match: 12.71% Indels: 57
DB: 2 Gaps: 9
US-09-423-575-1 (1-1886) x C84670 (1-431)
Qy 841 GACCAAGCTCTTCGACGACGACAAAGACATTTCTAATAACGTTGTTACTCAAGTTTCTCT 900
Db 60 GluGlnGlyAsnSerSerIleSerThrPheSerAsnGlyGlyValPheArgAlaLeuAla 79
Qy 901 CAACTTATATTGGCTCAAAATACCTTCACTCTGTTTCAAGAAATACTA----- 948
Db 80 ProIle-----TyrLeuLysAlaAlaGlnGluLeuMetVal 94
Qy 949 -----TCTCATTTGCGCGATACCTCTGCTGATTTTCAAGATATTAAGTCTCTT 1053
Db 95 AsnValGlyAsnGlySerHisGlyAlaLysGlnGluArgProValSerLysGluSerThr 114
Qy 994 GAGTCAGGAGCTCTAGTTCAGCTTACTTCAAGTTTTCAGATATTAAGTCTCTT 1053
Db 115 IleTyrGly-----ValGluAspIle 121
Qy 1054 GATGGTGATTTCTAATAACTCGGCGGGGTTTCGGATCTACATTTCAAGAGGAGATTA 1113
Db 122 AsnGlyGlyTyrLysProGlyValAla-----AlaLeu 132
Qy 1114 GAAGCAAGAAACCATCTCTTGGATCTCTTCAATGTTGGATGATGATATAGTAT 1173
Db 133 GlnMetLysLysAlaLysLeuIleSerMetGlyGluMetValGluGlnArgTyrLysGln 152

Qy 1174 TCGGTAGATGAGATTCATACGGTTTATATCAGCGTTTCCATGCTGCAACCGAGTTA---GAT 1230
Db 153 TyrHisAspGlnMetGlnThrIleIleSerPheGlnGlnAlaAlaGlyLeuGlySer 172
Qy 1231 CCACAGTTTACACACCGGTTTCCCTCCAAACCGTTTCTCTTATACAAAGACCTGAGA 1290
Db 173 AlaAsnSerTyrThrHisMetAlaLeuGlnThrIleSerLysGlnPheArgAlaValLys 192
Qy 1291 GAGAAATCTCAAGAGAGATATCTCTATGGGATCTGTATTGGAGAGAGCAAGACAAAG 1350
Db 193 AspMetIleSerLeuGlnIleLysGlnIleAsnLysLeuLeu-----GlyGlnLysGlu 210
Qy 1351 ACTCAAGAAACC-----TCTATGTTCCACGACCATGTCCTTCTTCAGCAG 1395
Db 211 PheAspGluGlnLeuLysLysLeuGlyLysMetAlaHisHis----- 225
Qy 1396 CTGAAACGAAAGAACCATCATGATTGGAGACCTCAACGAGGTTTCCCTGAGAAATCTGTT 1455
Db 226 -----SerAsnAlaTrpArgProGlnArgGlyLeuProGluLysAlaVal 240
Qy 1456 TCGGTTTACGGAATTTGGATCTTCAAACTTCTTCCCTTACCCGAAAGATTCGGAG 1515
Db 241 SerValLeuArgSerTrpLeuPheGluHisPheLeuHisProTyrProArgAspLeuAsp 260
Qy 1516 AAACATCTCTAGCTATACGAAGTGGCTTGACAAAGATCAGGATCAAACTGGTTTATA 1575
Db 261 LysValMetLeuAlaLysGlnThrGlyLeuThrLysSerGlnValSerAsnTrpPheIle 280
Qy 1576 AATCGCGGTTTAGCTATGGAACCGATGATAGAAAGATGTATGCGGAA----- 1626
Db 281 AsnAlaArgValArgMetTrpLysProLeuValGluGluLeuTyrSerGluGluMetAsp 300
Qy 1627 ATGAACAGAGGAAGCTCAATAACAGTCACATTAACCCACGACCACTTTCGGAATG 1686
Db 301 IleGluGluSerArgLysGlySerAspArgTyrSerThrLysGlySerSerLysGln 320
Qy 1687 CCA 1689
Db 321 Pro 321
RESULT 8
T05281
Probable homeobox protein T4L20.190 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05281
R.; Bevan, M.; Terry, N.; Ardiles, W.; Buysschaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, September 1998
A:Reference number: Z15406
A:Accession: T05281
A:Molecule type: DNA
A:Residues: 1-532 <BEV>
A:Cross-references: EMBL:AL023094
A:Experimental source: cultivar Columbia; BAC clone T4L20
C:Genetics:
A:Map position: 4
A:Introns: 217/3; 340/2; 360/3
A:Note: T4L20.190
Alignment Scores:
Pred. No.: 5,24e-25 Length: 532
Score: 417.50 Matches: 134
Percent Similarity: 44.52% Conservative: 61
Best Local Similarity: 30.59% Mismatches: 159
Query Match: 12.71% Indels: 85
DB: 2 Gaps: 12
US-09-423-575-1 (1-1886) x T05281 (1-532)
Qy 481 AATTCTATGCTGACTAACAATACTTATCTTCTCCACGGTCTATCGATGTTCAAGAT 540
Db 13 AspSerMetIleGlnAsnAlaIleValSerTyrSerGluGluSerAlaGly----- 29

Db 146 GlnGlnGlnGlyTyrSerGlyAsnLysSerThrGlnHisGlnAsnLeuGlnHisThrGln 165
 QY 829 AGATTAGCCTCAGACCAAGCTCTTTCAGCAGCAAGACATTTCTAATAACGTTCTTACT 888
 Db 166 MetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMet 185
 QY 889 CAAGGTTCTCTCAACTTATATTGGC---TCAAAATACCTTCACTCTCTTCAAGAAATA 945
 Db 186 HisAsnHisGlnPheGlnGlySerSerLysTyrLeuSerProAlaGlnGluLeu 205
 QY 946 CTATCTCATTTCCGCCACTACTCGCTCGATTATTCATCT----- 984
 Db 206 LeuSerGluPheCysSerLeuGlyValLysGluSerAspGluValMetMetMetLys 225
 QY 985 ---CGAGGACACCGAGTCAGAGCTGTAGTTCAGCCCTTT---ACTTCAGCTTTTTCAGAT 1038
 Db 226 HisLysLysLysGlnGlyGlnGlnGluGluTyrAspThrSerHisHisSerAsn 245
 QY 1039 ATAACTGAGTTCTTATGGTGTATCTTAATACTCGAGCGCGGTTTCGATCT----- 1092
 Db 246 -----AsnAspGlnHisAspGlnSerAlaThrThrSerSerLysLys 259
 QY 1093 -----ACATTTCAAAGGAGAGCATTTAGAAGCAAGAAAGAAACCCATCTC 1134
 Db 260 HisValProLeuHisSerLeuGluPheMetGluLeuGlnLysArgLysAlaLysLeu 279
 QY 1135 TTGGATCTCTTCAAATGTGGATGATCGATATAGTCATTGCGGTAGATGATTCATACG 1194
 Db 280 LeuSerMetLeuGlnLysLysArgArgTyrGlyHisTyrArgGluGlnMetArgVal 299
 QY 1195 GTTATATACGCTTCCATCTGCAACCGAGTTA---GATCCACAGTTACACACCGGTTT 1251
 Db 300 AlaAlaAlaPheGluAlaValGlyLeuGlyAlaGluLysThrAlaLeu 319
 QY 1252 GCCTCCAAACCGTTTCTCTTATCAAGAACCTCGAGAGAGATCTGCAAGAGATA 1311
 Db 320 AlaSerArgAlaMetSerArgHisPheArgCysLeuLysAspGlyLeuValGlyGlnIle 339
 QY 1312 ATCTCTATGGATCTGTATTG---GAGAGAGGCAAGAC----- 1347
 Db 340 GlnAlaThrSerGlnAlaLeuGlyGluArgGluGluAspAsnArgAlaValSerIleAla 359
 QY 1348 -----AAGACTCAAGAAACCTCTATGTTCCACGACGATTCCTTCCAGACGTTG 1398
 Db 360 AlaArgGlyGluThrProArgLeuArgLeuLeuAspGlnAlaLeuArgGlnLysSer 379
 QY 1399 AAACGAAAGAACCATCAGATT-----TGGAGACCTCAACAGAGTTTGCCTGAG 1446
 Db 380 TyrArgGlnMetThrLeuValAlaHisProTyrArgProGlnArgGlyLeuProGlu 399
 QY 1447 AAATCTGTTTGGCTTCTACGGAATTCGATGTTCCAAACTTCTTCCACCTTACCCGAA 1506
 Db 400 ArgAlaValThrThrLeuArgAlaThrLeuPheGluHisPheLeuHisProTyrProSer 419
 QY 1507 GATTCGGAGAAACATCTTAGCTATACGAAGTGGCTTGACAAGAGTCAGGTATCAAC 1566
 Db 420 AspValAspLysHisIleLeuAlaArgGlnThrGlyLeuSerArgSerGlnValSerAsn 439
 QY 1567 TGGTTTATAATGCGCGGTTAGCTATGAGAGCCGATGATACAGAGATGTATCGGAA 1626
 Db 440 TrpPheIleAsnAlaArgValArgLeuTyrPheProMetIleGluMetTyrCysGlu 459
 QY 1627 ATGAACAAGAGGAAGTCAATACAGTCACTTCAACCA----- 1666
 Db 460 Glu-ThrArgSerGluGlnMetGluIleThr---AsnProMetMetIleAspThrLysPr 478
 QY 1667 ----ACGGACCAACTCTTCGATGCCAA 1690
 Db 478 cAspProAspGlnLeuIleArgValGlu 487

RESULT 10

T02415

probable homeodomain transcription factor [imported] - Arabidopsis thaliana

N:Alternate names: probable BEL1-like homeotic protein
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 16-Feb-2001
 R:Accession: T02415; F84628
 R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, L.
 submitted to the EMBL Data Library, May 1998
 A:Description: Arabidopsis thaliana chromosome II BAC F27L4 genomic sequence.
 A:Reference number: 214658
 A:Accession: T02415
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-627 <R0U>
 A:Cross-references: EMBL:AC004482; NID:g3152602; PIDN:AAC17087.1; PID:g3152608
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, L.
 Nature 402, 763-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: F84628
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-627 <STO>
 A:Cross-references: GB:A8002093; NID:g3152608; PIDN:AAC17087.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: F27L4.6; At2g23760
 A:Map position: 2
 A:Introns: 319/3; 450/2; 470/3

Alignment Scores:

Pred. No.: 3,05e-24 Length: 627
 Score: 408.00 Matches: 142
 Percent Similarity: 41.48% Conservative: 82
 Best Local Similarity: 26.30% Mismatches: 141
 Query Match: 12.42% Indels: 176
 Ds: 2 Gaps: 18

US-09-423-575-1 (1-1886) x T02415 (1-627)

QY 310 AACAAACAACAACAACACTTTTAGTTCTCTGGATAATGTCATGACTAACCAAAATCCT 369
 Db 93 AsnAsnAsnAsnSerThrLeuHisMetLeu----- 103
 QY 370 CTTCTCATGATTTTATACCTTCAAGAGAAGATTCAACTTCATTC-----TCAACA 420
 Db 104 -----LeuProAsnHisGlnGlyPheAlaPheThrAspGluAsnThr 118
 QY 421 ATGCTTCCA-----TGGAAATACCATCATGATCTCTCTACAA 459
 Db 119 MetGlnProGlnGlnGlnHisPheThrTyrProSerSerSerSerSerHisGln 138
 QY 460 -----ATGGGTGGCTTGATATTTTCAATTCTATGCTGACTACAAATACTTA 507
 Db 139 AsnArgAspMetIleGlyThrValHisValGluGlyLysGlyLeuSerLeu 158
 QY 508 TCATCTTCTCCACGGTCTATCGATGTTCAAGATAACCGCAATGTTGA-----GTT 557
 Db 159 SerSerSerLeuAlaAlaAlaLysAlaGluGluTyrArgSerIleTyrCysAlaAlaVal 178
 QY 558 CATGGCTCTCTCTCTCATCTCTCTCCACTTCA-----TCCTTTTGATCATTTAAG 608
 Db 179 AspGlyThrSerSerSerSerAsnAlaSerAlaHisHisGlnPheAsnGlnPheLys 198
 QY 609 ACATCATGATGATTCCTCAACAA----- 632
 Db 199 AsnLeu-----LeuLeuGluAsnSerSerSerGlnHisHisGlnValValGly 216
 QY 633 CATGTGGGGTTTTGAAGCAAAATAGTGTTCAGGCATTTTCAGGTGTAGTTGTTCAAG 692
 Db 217 His-PheGlySerSerSerSerProMetAlaAlaSerSerSerIleGlyGly----- 234
 QY 593 TGAACCAATGATGCTCATATTCGATGAGAGATTTCCCGTTTCTTAATTCGAATAAAG 752

Db 235 -----flerYrThrLeuAr 239
 QY 753 AAACAATGAGCTTTCATTGAGCTTGTGCATCAGATGTTCTTCATGATGATCGGAGATAAG 812
 Db 239 GAsn----- 240
 QY 813 TCTTTGTGCGAGCTACAGATTAGCTCCTCAGACAAAGCTTCTTGCAGCAGCAAAAGACATTTC 872
 Db 240 ----- 240
 QY 873 TAATAACGTTGTACTCAGGGTTCTCTCAACTATATTTGGCTCAAAATACCTTCACCTC 932
 Db 241 -----SerLysYrThrLysPr 246
 QY 933 TGTTCAGAAATACATCTCATTTCCCGCCATCTCGCTCGATTATTCATCTCGAGGAAC 992
 Db 246 oAlaGlnGluLeuGluGluPheCysSerValGlyArgGlyHisPheLysLysAsnLy 266
 QY 993 CGAGTCAGAGCTGTAGTTCAGCCTTTACTTCAGCTTTTCAGATATATACCTAGTTCT 1052
 Db 266 sLeuSerArgAsnAsnSerAsn-----ProAsnThrThrGlyGly 280
 QY 1053 TCATGCTGATTCTAATACTCGAGCGGGTTTCGGA-----TCTAC 1094
 Db 280 yGlyGlyGlySerSerSerSerAlaGlyThrAlaAsnAspSerProProLeuSerPr 300
 QY 1095 ATTCAAGAGAGCATTAGAAAGCAAAACCCATCTCTTGGATCTCTTCAATGGT 1154
 Db 300 oAlaAspArgIleGluHisGlnArgGlyValLysLeuLeuSerMetLeuGluGluVa 320
 QY 1155 GGTATGATCATATAGTCTGCTGATGATCAGATCATACGCTTATATCAGCTTCCATGC 1214
 Db 320 lAspArgArgYrAsnHisTyrCysGlnMetGlnMetValValAsnSerPheAspGl 340
 QY 1215 TCAACCGAGTTAGATCCACAGTTA---CACACCGGTTTCCCTCCAAACCGTTTCCCT 1271
 Db 340 nValMetGlyTyrglyAlaAlaValProTyrThrThrLeuAlaGlnLysAlaMetSerAr 360
 QY 1272 CTTATACAGACCTTGAGAGAGATCTGCAAGAGATATC----- 1314
 Db 360 gHisPheArgCysLeuLysAspAlaValAlaValGlnLeuLysArgSerCysGluLeuLe 380
 QY 1315 -----TCTATGCGATCTGTATTGAGAGAGGC----- 1341
 Db 380 uGlyAspLysGluAlaAlaGlyAlaAlaSerSerGlyLeuThrLysGlyGluThrProAr 400
 QY 1342 -----AAAGACAAGACTCAAGAAACCTCTATGTTCCACAGCATTCCTCTCT 1388
 Db 400 gLeuArgLeuLeuGluGlnSerLeuArgGlnGlnArgAlaPheHisMetGlyMetMe 420
 QY 1389 TCAGAGCTGAAACGAAACCAACCATCAGATTGGAGACTCAACAGAGCTTTCCTCGAGAA 1448
 Db 420 tGlnGln-----GluAlaTrpArgProGlnArgGlyLeuProGluAr 434
 QY 1449 ATCTGTTTCGTTCTACGAAATGGATGTTCAAACTCTCTTCCCTTCCCTTCCCGGAAAGA 1508
 Db 434 gSerValAsnIleLeuArgAlaTrpLeuPheGluHisPheLeuAsnProTyrProSerAs 454
 QY 1509 TTCGAGAAACATCTCTAGCTATACGAGTGGCTTGCAGAACTCAGGTATCAAACTG 1568
 Db 454 pAlaAspLysHisLeuLeuAlaArgGlnThrGlyLeuSerArgAsnGlnValSerAsnTr 474
 QY 1569 GTTTATAATCGCGGGTTAGCTATGGAGCCGATGATAGAGAGATGTAT----- 1620
 Db 474 pPheIleAsnAlaArgValArgLeuTrpLysProMetValGluGluMetTyrGlnGlnGl 494
 QY 1621 -----GCCGAATGACAGAGGAGCTCAATACAGTCATCA 1661
 Db 494 uAlaLysGluArgGluGluAlaGlu--GluGluAsnGluAsnGlnGlnGlnArgArg 513
 QY 1662 ACCACGAGCAACATCTCTCGAATGCCAAATCTCTGTTATGATGAGCCCAAGCAAT 1715
 Db 514 GlnGlnGlnThrAsnAsnAsnAspThrLys-----ProAsnAsn 526

RESULT 11

D84679
 Probable homeodomain transcription factor [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C;Accession: D84679
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; A.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.B.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A;Reference number: A84420; MUID:20083487; PMID:10617197
 A;Accession: D84679
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-584 <STO>
 A;Cross-references: GB:AE002093; NID:g4510417; PIDN:AAD21503.1; GSPDB:GN00139
 C;Genetics:
 A;Gene: At2g27990
 A;Map position: 2

Alignment Scores:
 Pred. No.: 1,08e-23 Length: 584
 Score: 401.00 Matches: 139
 Percent Similarity: 42.42% Conservative: 82
 Best Local Similarity: 26.68% Mismatches: 172
 Query Match: 12.21% Indels: 128
 DB: 2 Gaps: 19

US-09-423-575-1 (1-1886) x D84679 (1-584)

QY 250 AATCTTCCCAAGGTTCTATCAGAGAAGAGATTAAGTTTCATAGAAACCAATGGAC 309
 Db 70 AsnMetProGln-----SerIleHisArgAspProGlnGly 81
 QY 310 AACACAACAC----- 318
 Db 82 ProSerAsnTrpArgIleSerAspLeuSerGlnProSerThrValAsnHisGlyTyrAsp 101
 QY 319 -----AACACAACACTTTTAGTTCTCTCGATAATGTCATGACTAACCAA 363
 Db 102 GlnAlaGlyIleArgProAsnAsnValAlaAspLeuLeuSerAspHisPheSerSerArg 121
 QY 364 AATCTCTCTCTCATGAGTTTATACCTTCAAGAGAAGATTCAACTTCATCTCAACAATG 423
 Db 122 AsnGlnIleLeuAspArgProLeuTyrValGlyArgAspSerIleProGlnSerMet 141
 QY 424 CTTCCATGGAATACCATCAGATCA---GATCCTCTCAAAATGGTGGCTTTCATATTTTC 480
 Db 142 IleArgArgSerGluValSerCysLeuAspAspAsnGlnLysGlyCysValThrValAla 161
 QY 481 AATTCTATGCTGACTAACAAATACCTTATCATCTTCTCCACGGTCTCTCATGATGTTCAAGAT 540
 Db 162 CysSerGlyThrGlyAsnGluIleLeuArgSerSerTyrAspGlnGlySerSerGly 181
 QY 541 AACCGCAATGTTGAGTTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
 Db 182 SerTyrArgGlyGluPheSerPheLeuPro-----SerLeuGlu 194
 QY 601 CATTTAAGACACATCATGATCTCTCAAAACACATCTGGGGTTTGAAGCAAAATAGTAG 660
 Db 195 AsnGlnSerValAlaHisAsnAlaSerAsn-----TrpAsnHis----- 207
 QY 661 TTTTCAGGCATTTTCAGGTGTAGTGTGTCAGAGTGAACCAATGATGTCT-----ACATTC 714
 Db 208 -----GlyProValAsnValThrAlaThrSerHisThrAsn 219
 QY 715 GGTGAAGAGATTTCCCGTTTCTTAATTTCCGATTAAGAGAAACAAATGAGCTTTTCATTGAGT 774
 Db 220 SerLysLysGlyPhePro-----LeuSer 227
 QY 775 CTTGTCATCAGATGTTTCTGATGAATCTCGAGATAAGTCTTTTGTGTCAGCTCAAGATTA 834


```

Db      228 LeuLeuSerAspIlePro----- 233
QY      835 GCCTCAGACAGCTTCTTGCACGACAAAGACATTTCTAATAACGTTTACTCAA----- 891
Db      234 -----ProSerArgAspValGlyAsnAlaValLeuSerThr 246
QY      892 -----GGTTTCTCAACTTATATTTGGCTCA 918
Db      247 MetAsnIleHisGlyProLeuGlyProPheThrGlyTrpAlaSerIleLeuLysSerSer 266
QY      919 AATACCTCTCACTCTCTCAAGAAATACATCTCATTTCCGCGCATACCTCGCTCGATTAT 978
Db      267 ArgPheLeuGluProAlaGlnLysMetLeuGluGluPheCys-----IleSerTyr 283
QY      979 TCATCTCGAGGACCGAGTCGAGAGCTGTAGTTACGCTTTACTTACGTTTTCAGAAAT 1038
Db      284 AlaSerLys-----IleIleSerArgSerGluSer 293
QY      1039 ATAACCTGAGTTTCTTGATGGTGATCTTAATACTCGAGGCGGGTTTCGATCTACATTT 1098
Db      294 ThrSerMetGluAspAspAspAspAspAsnLeuSerGlyPheSerSerSerSer 313
QY      1099 CAAGGAGAGCATTAGAA-----GCAAGAAACCCATCTCTTGGATCTTTCT 1146
Db      314 Glu-----ProLeuGluProLysAsnArgLeuLysLysAlaLysLeuLeuPheLeuGln 331
QY      1147 CAATGGTGGATGATCGATATAGTCATTCGCTAGATGAGATTCATACGGTTATATACGCG 1206
Db      332 GluGluValCysLysTrpTyrLysLeuTyrAsnHisGlnLeuGlnThrValMetSerSer 351
QY      1207 TTCCATGCTGCAACCGAGTTAGTACACAGTTA---CACACCGGTTTGGCTTCCAAACC 1263
Db      352 PheAsnThrValAlaGlyLeuAsnThrAlaThrProTyrIleSerLeuAlaLeuLysArg 371
QY      1264 GTTTCCTCTTATACAGAACCTGAGAGAGAACTCCAGAGAACTCTATGTTCCACGAGCATTGC 1383
Db      372 ThrSerArgSerPheLysAlaLeuArgThrAlaIleAlaGluHisValLysGlnIleSer 391
QY      1324 TCTGTATTGGAGAGAGCAAGACAGACTCAAGAAACCTCTATGTTCCACGAGCATTGC 1383
Db      392 SerHisSerSerAsnGlyAsnAsnAsnAsnArgPheGlnLys-----ArgGlnArgSer 409
QY      1384 CTTCTTCACGAG-----CTGAACGAAAGAACCATTCAGTTTGAGACCTCAACGA 1434
Db      410 LeuIleGlyAsnAsnValGlyPheGluSerGlnGlnGlnHisIleTrpArgProGlnArg 429
QY      1435 GGTTCCTGCTGAGAAATCTGTTTCGTTCTACGAAATGATGTTTCCAAACTTCTTCCAC 1494
Db      430 GlyLeuProGluArgAlaValAlaValLeuArgAlaTrpLeuPheAspHisPheLeuHis 449
QY      1495 CTTTACCGAAGATTCGAGAAACATCTTCTAGCTATACGAAGTGGCTTACCAAGAAGT 1554
Db      450 ProTyrProThrAspSerAspLysGlnMetLeuAlaThrGlnThrGlyLeuSerArgAsn 469
QY      1555 CAGGTATCAACTGTTTATAATCGCGGTTAGGCTATGAGCCGATCATAGACAG 1614
Db      470 GlnValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetValGluGlu 489
QY      1615 ATGTATCGGAAATGAACAAG-----AGGAAGCTCAATAACAGTCAAC---ATTCAACCC 1665
Db      490 IleHisThrLeuGluThrLysAlaIleLysAsnAlaAspThrSerHisAsnIleGluPro 509
QY      1666 AAC 1668
Db      510 Ser 510

```

RESULT 12

T48224

probable homeodomain protein - Arabidopsis thaliana

N/Alternate names: protein T7H20.80

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C/Accession: T48224
 R/Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew
 submitted to the Protein Sequence Database, March 2000
 A/Reference number: 224488
 A/Accession: T48224
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-575 <BEV>
 A/Cross-references: EMBL:AL162508
 A/Experimental source: cultivar Columbia; BAC clone T7H20
 C/Genetics:
 A/Map position: 5
 A/Intons: 240/3; 374/2; 394/3
 A/Note: T7H20.80

Alignment Scores:

Pred. No.: 4.95e-22 Length: 575
 Score: 380.00 Matches: 126
 Percent Similarity: 40.72% Conservative: 67
 Best Local Similarity: 26.58% Mismatches: 165
 Query Match: 11.57% Indels: 116
 DB: 2 Gaps: 14

US-09-423-575-1 (1-1886) x T48224 (1-575)

```

QY      532 GTTCAAGATAACCGCAATGTTGAGTTTCATGCTCCT----- 567
Db      11 LeuGlnGlnSerArgArgAspLysLeuArgIleProSerLeuAspSerHisPheHisPhe 30
QY      568 ---CCTCCTCATCCTCCT-----CCACTTCATCCTTTG 597
Db      31 HisProProProProSerSerGlyGlyGlyGlyValPheProLeuAlaAspSer 50
QY      598 GATCATTTAAGACACATGATGATTCCTCAACAACATGTTGGGTTTGAAGCAATAGT 657
Db      51 AspPheLeuAlaGlyGlyPheHisSerAsnAsnAsnHisIleSerAsnPro 70
QY      658 GAGTTTCAGGCATTTTCAGGTGTAGTTGTT-----CCAAGTGAACCAATCATG 705
Db      71 SerTyrSerAsnPheMetGlyPheLeuGlyGlyProSerSerSerSerThrAlaVal 90
QY      706 TCTACATTCGTTGAAGAGATTTCCCGTTTCTAATTCGAATAAAGAAACAATGAGCTT 765
Db      91 AlaValAlaGlyAspHisSerPhe-----AsnAlaGlyLeu 102
QY      766 TCATTGAGTCTTCATCAGATGTTTCTGATGAATGTCGAGAGATAAGTCTTTGTCAGCT 825
Db      103 SerSerGlyAspValLeuValPheLysProGluProLeuSerLeuSerSerHis 122
QY      826 ACAAGATTAGCCTCAGAG-----CAAGCTTCTTTCAGCAGC 861
Db      123 ProArgLeuAlaTyrAspLeuValValProGlyValValAsnSerGlyPheCysArgSer 142
QY      862 AAAGACATTTCTAATAACGTTGTACT----- 888
Db      143 AlaGlyClyAlaAsnAlaAlaValThrIleAlaSerArgSerSerGlyProLeuGly 162
QY      889 -----CAAGGTTTCTCTCAACTTATATTGCTCAAAATACCTTCACTCTGTCAAGAA 942
Db      163 ProPheThrGlyTyrAlaSerIleLeuLysGlySerArgPheLeuLysProAlaGlnMet 182
QY      943 ATACTATCTCAITTCGCCGCATACCTCGCTCGAT---TATTCACTCTCGAAGAACCGAGTCA 999
Db      183 LeuAspGluPheCysAsnValGlyArgGlyIleTyrThrAspLysValIleAspAsp 202
QY      1000 GGAGCTGTAGTTTCAGCCTTTACTTCACGTTTTTGAAGATAATTAAGTTCCTTGTATGTT 1059
Db      203 AspAspSerSerLeuLeuPheAspProThrValGluAsnLeuLysGlyValSerAspGly 222
QY      1060 GATCTTAATAACTCGGAGCGGGTTTCGGATCTACATTTCAAAGGAGAGCATTAGAAGCA 1119
Db      223 GlyGlyGlyAspAsn-----GlyLys 229

```


C;Comment: This protein, functions as a nuclear transcription factor that belongs to the and interacts with P53 proteins and plays a tissue-specific regulation of transcription.
A;Gene: PKnox2
A;Map position: chromosome 11q24

Alignment Scores:
Pred. No.: 9,93e-07 Length: 460
Score: 186.50 Matches: 82
Percent Similarity: 42.58% Conservative: 70
Best Local Similarity: 22.97% Mismatches: 131
Query Match: 5.68% Indels: 75
DB: 2 Gaps: 14

US-09-423-575-1 (1-1886) x JC7766 (1-460)

QY	811	AGCTTTGTCACGATACAGATTAGCTTCAGACAGCTTCTTGCAGACCAAGACATT	970
Db	26	SerLysAlaGlnAlaValHisIleSerAlaProSerAlaAlaAlaSerThrProValPro	45
QY	971	TCTAATAAACCTGTTACTACTAGGTTTCTCTCACTTATATTTGGCTCAAAATACCTTCAC	930
Db	46	SerAlaProIleAspProGlnAlaGlnLeuGluAlaAspLysArgAlaValTyrArgHis	65
QY	931	TCTGTTCAAGAAATATCTATCTATCTGCGGCGCATCTCGCTCGAATTATTCATCTCGAGGA	990
Db	66	ProLeuPheProLeuLeuThr---LeuLeuPheGluLysCysGluGlnAlaThrGlnGly	84
QY	991	ACCGAGTCAGGAGCTGCTAGTTCAGCTTACTTTCAGCTTTTGAGAAATACGAGTTT	1050
Db	85	SerGluCys---IleThrSerAlaSerPheAspValAspIleGluAsn-----Phe	100
QY	1051	CTTGATGGTGAATCTAATAACTCGAGCGGGGTTTCGGA-----TCTACA	1095
Db	101	ValHisGlnGlnGlnGlnGlnHisLysProPhePheSerAspAspProGluLeuAspAsn	120
QY	1096	TTTCAAGAGAGACCATAGACGAAGAAACCATCTCTTGGATCTTCTTCAATGATGG	1155
Db	121	LeuMetValIleAlaIleGlnValLeuArgGileHisLeuLeuGlu---LeuGluLysVal	139
QY	1156	GATGATCGATATAGTCATGCGTAGATGAGATTTCATACGGTTATATCATCGCGTTC	1209
Db	140	AsnGluLeu-----CysLysAspPheCysAsnArgTyrIleThrCysPheLysThr	156
QY	1210	-----CATGCTGC-AACGAGTTAGATCCACAGTTTACACCCCGTTTGCCT-----	1256
Db	157	LysMetHisSerAspAsnLeuLeuArgAsnAspLeuGlyGlyProTyrSerProAsnGln	176
QY	1257	-----CCAAACCGTTCTCTTATACAGAACCTCTGAGAGAGAGATCTG	1301
Db	177	ProSerIleAsnLeuHisSerGlnAspLeuLeuGlnAsnSerProAsnSerMetSerGly	196
QY	1302	CAAGAGATAATCTCTATGGGATCTGTATTGGAGAGAGGCAAGACAGAC-----	1352
Db	197	ValSerAsnAsnProGlnGlyIleValValProAlaSerAlaLeuGlnGlnGlyAsnIle	216
QY	1352	-----	1352
Db	217	AlaMetThrThrValAsnSerGlnValValSerGlyAlaLeuTyrGlnProValThr	236
QY	1353	-----TCAAGAAACCTCTATGTTCCACGAGATTCGCTTCTTCAGCAGCTGAAACG	1403
Db	237	MetValThrSerGlnGlyGlnValValThrGlnAlaIleProGlnGlyAla-IleGlnIle	256
QY	1404	AAAGAACCATTCAGAT-----	1421
Db	256	eGlnAsnThrGlnValAsnLeuAspLeuThrSerLeuLeuAspAsnGluAspLysLys	276
QY	1422	GAGACCTCAACGAGGT---TTCCCTGAGAAATCTGTTTCGGTCTTACCGAATTCGATGTT	1478
Db	276	rlYsnLysArgGlyValLeuProCysHisAlaThrAsnIleMetArgSerTrpLeuPh	296
QY	1479	CCAAACACTTCCTTCACCTTACCCGAAAGATTTCGGAGAAACATCTCTTAGCTATACGAAG	1538

RESULT 15

S42543
hypotheical protein - soybean
C;Species: Glycine max (soybean)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Sep-1999
C;Accession: S42543
R;Ma, H.; McMullen, M.D.; Finer, J.J.
Plant Mol. Biol. 24, 465-473, 1994
A;Title: Identification of a homeobox-containing gene with enhanced expression during so
A;Reference number: S42543; MUID:94169300; PMID:7907232
A;Accession: S42543
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-379 <MAH>
A;Cross-references: EMBL:L13663; NID:g310568; PID:AAA20882.1; PID:g485406
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;284-343/Domain: homeobox homology <HOX>

Alignment Scores:
Pred. No.: 5,42e-06 Length: 379
Score: 177.00 Matches: 92
Percent Similarity: 35.31% Conservative: 69
Best Local Similarity: 20.18% Mismatches: 141
Query Match: 5.39% Indels: 154
DB: 2 Gaps: 20

US-09-423-575-1 (1-1886) x S42543 (1-379)

QY	298	AACCAATGGACACAAACAAACACATTTTATGTTCTCTGGATAATGTCATGACT	357
Db	42	HisProIleAsnProSerAsnAsnAsn-----Val	52
QY	358	AACCAAAATCTCTCTCTATGAGTTTATACCTTCAAGAGAGATTCACATTCATCTCA	417
Db	53	AsnThrAsnCysLeu-----PheIleProAsnCysSerAsnSerThrGlyThrPro	69
QY	418	ACATGCTTCCATGGAATACCATCAGATCAGATCCCTCTACAAATGGTGGCTTTGATTT	477
Db	70	SerIleMet-----Leu	73
QY	478	TTCAATTTATGCTGACTAAACAAATATCTATCTCTCCAGGCTCTATCGATGTTCAA	537
Db	74	HisAsnAsnHisAsnAsnLys-----ThrAspAspAsp	85
QY	538	GATAACCGCAAT-----GTTGAGTTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCT	582
Db	86	AspAsnAsnAsnAsnThrGlyLeuGlyTyrTyrPheMetGluSer-----	100
QY	583	CCACTTTCATCTTGGATCATTTAAGACACTATGATGATCTCTCCAAACACATGTGGGT	642
Db	101	-----AspHisHisHisHisHisHisHisGlyAsnAsnAsnAsnAsn-----Gly	114
QY	643	TTTGAAGCAAAATAGTGGTTTCAGGCAATTTTCAGGTGTAGTTGGTCCAAAGTGAACCAATG	702
Db	115	Ser	128
QY	703	ATGCTACATTCGGTGAAGAGATTTCCTGTTTCTAATTCGATATAAAGAACAAATGAG	762

Thu Sep 2 15:34:08 2004

```

Db 129 -----MetAlaHisProHisTyrHisArgLeuLeuAlaAlaTyrValAsnCysGln 145
QY 763 CTTTCATTGAGCTTCATGATGTTTCTGATGATGCTCGGAGATAAGTCTTTGTGCA 822
Db 146 ---LysValGlyAlaProGluValAlaArgLeuGluGluAlaCysAlaSerAla 164
QY 823 GCTCAAGATTAGCTCAGCAAGCT-----TCTTCAGCAGCAAGACATTCT 873
Db 165 AlaThrMetAlaGlyGlyAspAlaAlaAlaGlySerSerCysIleGlyGluAspProAla 184
QY 874 AATAAGCTTGTACTCAAGGTTCTCTCAACTATATTTGGCTCAAAATACCTTCACTCT 933
Db 185 LeuAspGlnPheMetGluAlaTyrCysGluMetLeu-----ThrLysTyrGluGlnGlu 202
QY 934 GTTCAAGAAATACTATCTCATTTCCCGCATACTCGCTCGATTATTCATCTCGAGGAACC 993
Db 203 LeuSerLysProLeuLys----- 208
QY 994 GAGTCAGGAGCTGCTAGTTACGCTTTTACGCTTTTGTGAGATATATACTGAGTTCTT 1053
Db 209 GluAlaMetLeuPheLeuGlnArgIleGluCysGlnPheLysAsnLeuThr---IleSer 227
QY 1054 GATGTGATTTCTAATACTCGGAGCGGTTTCGGATCTACATTTCAAAGGAGAGCATTA 1113
Db 228 SerSerAspPheAlaSerAsnGluGly----- 237
QY 1114 GAAGCAAGAAACCCATCTCTTGATCTTCTCAATGGTGGATCGATATAGTCAT 1173
Db 238 -----AspArgAsnGlySer 242
QY 1174 TCGGTAGATGAGATT-----CATACGGTTATATCAGCGTTTCCATGCTGCAACCGAGTTA 1227
Db 243 SerGluGluAspValAspLeuHisAsnMetIle----- 253
QY 1228 GATCCACAGTTACACACCGGTTGGCCCTCCAAACCGTTTCTCTTATACAGAACCTG 1287
Db 254 AspProGlnAlaGluAspArg-----AspLeu 262
QY 1288 AGAGAGAGAAATCGCAAGAGATAATCTCT---ATGGGATCTGTATTGGAGAGAGGCAAA 1344
Db 263 LysGlyGlnLeuLeuArgLysTyrSerGlyTyrLeuGlySer----- 276
QY 1345 GACAAGACTCAAGAAACCTCTATGTTCCACAGCATTCGCTTCTCAGCAGCTGAAACGA 1404
Db 277 -----LeuLysGlnGluPheMetLys 283
QY 1405 AGAACCATCAGATTGGAGACCTCAACGAGGTTTGCTTGAGAAATCTGTTTCGGTTCTA 1464
Db 284 Lys-----ArgLysLysGlyLysLeuProLysGluAlaArgGlnGlnLeu 298
QY 1465 CGGAATTGGATGTTCAAAACTTCTTACCTTACCCGAAAGATTCGGAGAAACATCTT 1524
Db 299 LeuGluTyrTrpAsnArgHisTyrLysTrpProTyrProSerGluSerGlnLysLeuAla 318
QY 1525 CTAGCTATACGAGTGGCTTGACAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1584
Db 319 LeuAlaGluSerThrGlyLeuAspGlnLysGlnIleAsnAsnTrpPheIleAsnGlnArg 338
QY 1585 GTTAGGCTATGGAAGCCGATGATAGAGAGATGATGCGGAAATGAAC 1632
Db 339 LysArgHisTrpLysProSerGluAspMetGlnPheValMetAsp 354

```

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 2, 2004, 09:48:42 ; Search time 22 Seconds
(without alignments)
8927.665 Million cell updates/sec

Title: US-09-423-575-1
Perfect score: 3284
Sequence: 1 atttggtataaaattgtgc.....ttaaaaaaaaaaaaaaaaaaaaaa 1886

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US09423575/runat_01092004_161106_18556/app_query.fasta_1.2055
-DB=SwissProt_42 -QFMT=fascan -SUFFIX=rsf -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09423575 @CNG 1.1 26 @runat_01092004_161106_18556 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2484	75.6	473	ATH1_ARATH	P48731 arabidopsis
2	208	6.3	435	PNX1_MOUSE	O70477 mus musculus
3	202	6.2	435	PNX1_HUMAN	P55347 homo sapien
4	196.5	6.0	390	ME11_MOUSE	O60954 mus musculus
5	196.5	6.0	477	ME12_HUMAN	O14770 homo sapien
6	196.5	6.0	477	ME12_MOUSE	P97367 mus musculus
7	195.5	6.0	390	ME11_HUMAN	O00470 homo sapien
8	194	5.9	378	ME13_MOUSE	P97368 mus musculus
9	189.5	5.8	390	ME11_XENLA	P79937 xenopus lae
10	186.5	5.7	460	PNX2_HUMAN	O96kn3 homo sapien
11	182	5.5	382	ME13_HUMAN	O99687 homo sapien
12	177	5.4	375	HMB1_SOYBN	P46608 glycine max
13	175.5	5.3	355	HK16_LYCES	O22289 lycopersico
14	175.5	5.3	431	HK13_ARATH	P48000 arabidopsis
15	169.5	5.2	364	KNX3_HORVU	O43484 hordeum vul
16	169	5.1	291	SIX2_HUMAN	O99pc8 homo sapien
17	167	5.1	398	HK11_MALDO	O04134 malus domes
18	166.5	5.1	359	HKN1_MAIZE	P24345 zea mays (m

19	166.5	5.1	659	1	HMN1_DROME	P22807 drosophila
20	161.5	4.9	361	1	OSH1_ORYZA	P46609 oryza sativ
21	161.5	4.9	398	1	HKL1_ARATH	P46639 arabidopsis
22	158.5	4.8	419	1	HKL4_ARATH	P48001 arabidopsis
23	158	4.8	296	1	SIX2_MOUSE	O62232 mus musculu
24	157.5	4.8	269	1	AK3_CHICK	O30655 gallus gall
25	157.5	4.8	276	1	YGV6_YEAST	P33147 saccharomyc
26	154	4.7	1314	1	SWIL_YEAST	P09547 saccharomyc
27	153	4.7	294	1	HD1_BRANA	P46606 brassica na
28	152.5	4.6	306	1	CUP9_YEAST	P41817 saccharomyc
29	152.5	4.6	397	1	HKL2_MALDO	O04135 malus domes
30	150	4.6	272	1	TGIF_HUMAN	O15583 homo sapien
31	149.5	4.6	88	1	HKL3_MAIZE	P56661 zea mays (m
32	149.5	4.6	272	1	TGIF_MOUSE	P70284 mus musculu
33	149.5	4.6	355	1	HKN1_LYCES	Q41330 lycopersico
34	146	4.4	273	1	SIX1_MOUSE	O62231 mus musculu
35	146	4.4	284	1	SIX1_HUMAN	Q15475 homo sapien
36	145	4.4	383	1	HKL5_ARATH	P48002 arabidopsis
37	144	4.4	430	1	PEX1_HUMAN	P40424 homo sapien
38	144	4.4	430	1	PEX1_MOUSE	P41778 mus musculu
39	143.5	4.4	427	1	HKL3_MALDO	O04136 malus domes
40	143	4.4	329	1	HM40_CABEL	Q19503 caenorhabdi
41	142.5	4.3	351	1	RSH1_MAIZE	Q41853 zea mays (m
42	142	4.3	382	1	STM_ARATH	Q38874 arabidopsis
43	141.5	4.3	416	1	SO_DROME	O27350 drosophila
44	140.5	4.3	330	1	PEX4_HUMAN	O39yul homo sapien
45	140.5	4.3	426	1	HKLB_LYCES	O22300 lycopersico

ALIGNMENTS

RESULT 1

ATH1_ARATH
ID ATH1_ARATH STANDARD; PRT; 473 AA.

AC P48731;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homeobox protein ATH1
GN ATH1 OR AT4G32980 OR F26P21.100.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Seedling;
RX MEDLINE=95210927; PubMed=7696878;
RA Quaedvlieg N., Dockx J., Rook F., Weisbeek P.J., Smeekens S.;
RT "The homeobox gene ATH1 of Arabidopsis is derepressed in the
RL Plant Cell 7:117-129(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoerger W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hebeisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche P.,
RA Braeken M., Weljens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenecker T., Bothe G., Ransperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Mooijman P., Klein lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Baele H.,
RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin J., Quail M.A., Bray-Allen S.,

QY 1264 GTTTCCTCTTATACAGAACCTGAGAGAGAACTGCAAGAGATATCTCTATGGA 1323
 Db 321 ValSerPheLeuTyrLysAsnLeuArgGluArgIleCysLysIleSerMetGly 340
 QY 1324 TGTGTTATGAGAGAGCGCAAGACTCAAGAACTCTATGTTCCACGAGATTGC 1383
 Db 341 SerValLeuGluArgGlyLysAspLysThrGlnGluThrSerMetPheHisGlnHisCys 360
 QY 1384 CTTCTTCACAGCTGAAGCAAGAACCATCATGTTGGAGACCTCAACGAGTTTGCT 1443
 Db 361 LeuLeuGlnLeuLysArgLysAsnHisGlnIleTrpArgProGlnArgGlyLeuPro 380
 QY 1444 GAGAAATCTGTTTCGGTTCTACGGAATGCGATGTTCCAAAACCTCTCTCACCTTACCG 1503
 Db 381 GluLysSerValSerValLeuArgAsnTrpMetPheGlnAsnPheLeuHisProTyrPro 400
 QY 1504 AAGATTCGAGGAACATCTTAGCTATACGAGTGGCTGCACAGAGTCAAGTATCA 1563
 Db 401 LysAspSerGluLysHisLeuLeuAlaIleArgSerGlyLeuThrArgSerGlnValSer 420
 QY 1564 AACTGCTTTATAAATCGCGGGTTAGCTATGGAAGCCGATGATAGAAGAGATGTATCG 1623
 Db 421 AsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluMetTyrAla 440
 QY 1624 GAAATGACAGAGAGAGCTCAATAACAGTCACTCAATCAACCAACGAGCAACTCTTGA 1683
 Db 441 GluMetAsnLysArgLysLeuAsnAsnSerHisIleGlnProAsnGlyProThrLeuArg 460
 QY 1684 ATGCCAAAATCTGTTATGATGAGCCCAAGCAATGCATAAA 1722
 Db 461 MetProLysSerValMetMetSerGlnAlaMetHisLys 473

RESULT 2

PNK1 MOUSE
 ID PNK1 MOUSE STANDARD; PRT; 435 AA.
 AC Q70477;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Homeobox protein PNKX1 (PBX/knotted homeobox 1).
 GN PNKX1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss Webster;
 RA Nakamura T., Yamazaki Y.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: Belongs to the TALE/MEIS homeobox family.
 CC -1- SIMILARITY: Contains 1 homeobox domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF061270; AAC15990.1; -;
 DR HSSP; P40424; 1B72;
 DR TRANSFAC; T04121;
 DR MGD; MGI:1201409; Pknx1.
 DR GO; GO:0005667; C:transcription factor complex; IDA.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0003705; F:RNA polymerase II transcription factor acti. .; IDA.
 DR GO; GO:0006357; P:regulation of transcription from Pol II pro. .; IDA.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.

DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; Hox; 1.
 DR PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.
 DR PROSITE; PSS0071; HOMEBOX_2; 1.
 KW Homeobox; DNA-binding; Nuclear protein.
 FT DNA BIND 258 320 HOMEBOX (TALE-TYPE).
 SQ SEQUENCE 435 AA; 47395 MW; 354369A1795A1E84 CRC64;

Alignment Scores:

Pred. No.: 1,45e-08 Length: 435
 Score: 208.00 Matches: 89
 Percent Similarity: 41.53% Conservative: 63
 Best Local Similarity: 24.32% Mismatches: 122
 Query Match: 6.33% Indels: 92
 DB: 1 Gaps: 18

US-09-423-575-1 (1-1886) x PNK1_MOUSE (1-435)

QY 826 ACAAGATTAGCTCAGAGCAA---GCTTCTTGAGCAGCAAGAC----- 867
 Db 21 ThrGluLeuLysThrGluGlnAspProAsnCysSerAspProAlaGluGlyValSer 40
 QY 868 -----ATTCTAATAACGTTGTTACTCAAGGTTTCTCAACTATATATTTGGCTCA 918
 Db 41 ProProLleGluSerGlnThrProMetAspAlaAspLysGlnAlaIle----- 57
 QY 919 AAATACCTTCACTCTGTTCAGAAATACTATCTCATTTCCGCCGATACCTCGCTATAT 978
 Db 58 ---TyrArgHisProLeuPheProLeuAla---LeuLeuPheGluLysCysGluGln 75
 QY 979 TCATCTCGAGAACCGAGTCAGGAGCTGCTAGTGTTCAGCTTTTACCTTTCAGCTTTTGAAGAT 1038
 Db 76 SerThrGlnGlySerGlu---GlyThrSerAlaSerPheAspValAspIleGluAsn 94
 QY 1039 ATAAC-----GAGTTTCTGTGATGTTCTAATAACTCGGAGCGGGTTCCGATCT 1092
 Db 95 PheValArgLysGlnGluLysAspGlyLysProPheCysGluAspProGluThrAsp 114
 QY 1093 ACATTTCAAGGAGAGCATTAGAAGCAAGAAACCATCTCTTGGATCTTCTTCAATG 1152
 Db 115 AsnLeuMetValLysAlaIleGlnValLeuArgIleHisLeuLeuGluGluLysVal 134
 QY 1153 GTGGAT-----GATCGATATAGTCATTGCGTA----- 1179
 Db 135 AsnGluLeuCysLysAspPheCysSerArgTyrIleAlaCysLeuLysThrLysMetAsn 154
 QY 1180 -----GATGAG 1185
 Db 155 SerGluThrLeuLeuSerGlyGluProGlySerProTyrSerProValGlnSerGlnGln 174
 QY 1186 ATTCAATCGGTTATATCAGGTTCCATGCTGCAACCGAGTTAGATCCACAG-----TTA 1239
 Db 175 IleGlnSerAlaIleThrGly-----ThrLeuSerProGlnGlyIleVal 189
 QY 1240 CACACCCCGTTTGGCTTCCAA-----ACGTTTTC-----TTC 1272
 Db 190 ValProAlaSerAlaLeuGlnGlnGlyAsnValThrMetAlaThrValAlaGlyGlyThr 209
 QY 1273 TTATACAGAACCTG-----AGAGAGAGATCTGCAAGAGATATCTCT 1317
 Db 210 ValTyrGlnProValThrValValThrProGlnGlnValThrGlnAlaLeuSer 229
 QY 1318 ATGGGATCTGTATTGGAGAGAGGCAAA-----GACAAAGACTCAAGAACCTCT 1365
 Db 230 ProGlyThrIleArgIleGlnAsnSerGlnLeuGlnLeuAsnGlnAspLeuSer 249
 QY 1366 ATGTTCCACAGCATTCCTTCTCAGCAGCTGAACGAAAGAACCATCATGATTGGAGA 1425
 Db 250 ThrLeuHisGln-----GluAspGlySerSerLysAsn----- 260
 QY 1426 CCTCAACGAGGT---TTGCTCTGAGAAATCTGTTTCGGTTCTTACGGAATTGGATGTTCCAA 1482
 Db 261 ---LysArgGlyValLeuProLysHisAlaThrAsnValMetArgSerTrpLeuPheGln 279


```
QY 748 AAAAGAAACATGACGCTTTTCATGTAGTCTTGATCAGATGTT-----TCTGAT 795
Db 91 PheGluLysCysGluLeuAlaThrCysThrProArgGluProGlyValAlaGlyGlyAsp 110
QY 796 GAATGCTCGACATAAGTCTT-----TGTGACGCTCAAGATTAAGCTCT 840
Db 111 ValCysSerSerGluSerPheAsnGluAspLeuAlaValPheAlaLysGluLeuAla 130
QY 841 GAGCAAGCTTCTTGACGACGACAAACATCTTCTAATACGTTGTACTCAAGTTCTCT 900
Db 131 GluLysProLeuPheSerSerAsnProGluLeuAspAsnLeuMetIleGlnAlaIleGln 150
QY 901 CAACCTATATTGGCTCAAAATACCTCTCACTCTGTTCAGAAATACTATCTCATTTCC 957
Db 151 ValLeuArgPheHisLeuLeuGluLysValHisGluLeuLysCysAspAsnPheCys 170
QY 958 -----GCCGCATCTCGCTCGATTTATTCATCTCGAGGAACC 993
Db 171 HisArgTyrIleSerCysLeuLysGlyLysMetProIleAspLeuValIleAspArg 190
QY 994 GAGTCAGGAGCTGCTAGT---TCAGCCTTACTCTCACGTTTGTGAGATATACTAGTGT 1050
Db 191 GluGlyGlySerLysSerAspSerGluAspValThrArgSerAlaAsnLeuThr----- 208
QY 1051 CTTGATGTTGATTTAATAAATACTCGGAGCGGGTTTCGATCTACATTTCAAAGGAGAC 1110
Db 209 ---AspGlnProSerTrpAsn-Arg-----Asp-HisAspAspThrAlaSer 223
QY 1111 TTAGAAGAAAGAAACCCA-----TCTCTGGATCTCTCAATGGTGGATGAT 1161
Db 223 hrArgSerGlyGlyThrProGlyProSerSerGlyGlyHisThrSerHisSerGly-Asp 242
QY 1162 CGATATAGTCAUATCGGTAGATGATGATATACGGTTATATACGGTTCATCGCTCAACC 1221
Db 243 AsnSerSerGluGlnGlyAspGlyLeuAspAsnSerValAlaSerProSerThrGlyAsp 262
QY 1222 GAGTTAGATCCAGGTATACACACCGCGTTGCGCCCTCCCAACCGTTTCTCTTATACAG 1281
Db 263 AspAspAspProAsp----- 267
QY 1282 AACCTGAGAGAGAGATCTGCAAGAAGATATCTCTATGGGATCTGTATTGGAGAGAGCC 1341
Db 267 ----- 267
QY 1342 AAAGACAAGACTCAAGAAACCTCTATGTTCCACAGCATGTGCTTCTTCAGCAGCTGAAA 1401
Db 268 LysAspLys----- 270
QY 1402 CGAAGAAACATCAGATTTGGAGACCTCAACGAGTTTG---CCTGAGAAATCTGTTTCG 1458
Db 271 ---LysArgHisLys-----LysArgGlyIlePheProLysValAlaThrAsn 285
QY 1459 GTTCTACGGAATGGATGTTTCCAAAACCTTCCTTACCCCTTACCCGAAAGATTCGGAGAA 1518
Db 286 IleMetArgAlaTrpLeuPheGlnHisLeuThrHisProTyrProSerGluGluGlnLys 305
QY 1519 CATCTTCTAGCTATACGATGCTTGACAGAAAGTCAGGTATCAACTGTTTATAAAT 1578
Db 306 LysGlnLeuAlaGlnAspThrGlyLeuThrIleLeuGlnValAsnAsnTrpPheIleAsn 325
QY 1579 GCCCGGTTAGGCTATGGAAGCCGATGATAGAGAGATGATGTATCGGAATTAACAGAGG 1638
Db 326 AlaArgArgArgIleValGlnProMetIleAspGlnSerAsnArgAlaValSerGlnGly 345
QY 1639 AAGCTCAATACAGTCACATTCACCCACGGA 1671
Db 346 ThrProTyrAsnProAspGlyGlnProMetGly 356
RESULT 5
MEI2_HUMAN
ID MEI2_HUMAN STANDARD; PRT; 477 AA.
AC O14770; Q96KI4; Q96KI5; Q9NRS1; Q9NRS2; Q9NRS3;
```

```
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Homeobox protein Meis2 (Meis1-related protein 1).
GN MEIS2 OR MRGL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORMS MEIS2A; MEIS2B; MEIS2C; MEIS2D AND
RP MEIS2E).
RC TISSUE=Brain;
RX MEDLINE=20347166; PubMed=10764806;
RA Yang Y., Hwang C.K., D'Souza U.M., Lee S.-H., Junn E., Mouradian M.M.;
RT "Three-amino acid extension loop homeodomain proteins Meis2 and TGIF
RT differentially regulate transcription.";
RL J. Biol. Chem. 273:20734-20741(2000).
RN [2]
SEQUENCE FROM N.A. (ISOFORM MEIS2D).
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Sapletenko M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting J., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smaluk D.E.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
SEQUENCE OF 271-477 FROM N.A. (ISOFORM MEIS2B).
RX MEDLINE=98051942; PubMed=9383298;
RA Smith J.B. Jr., Afonja O., Yee H.T., Inghirami G., Takeshita K.;
RT "Chromosomal mapping to 15q14 and expression analysis of the human
RL MEIS2 homeobox gene.";
Mamm. Genome 8:951-952(1997).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Name=Meis2C;
CC IsoId=O14770-1; Sequence=Displayed;
CC Name=Meis2A;
CC IsoId=O14770-2; Sequence=VSP_002245, VSP_002246;
CC Name=Meis2B;
CC IsoId=O14770-3; Sequence=VSP_002242, VSP_002245, VSP_002246;
CC Name=Meis2D;
CC IsoId=O14770-4; Sequence=VSP_002242;
CC Name=Meis2E;
CC IsoId=O14770-5; Sequence=VSP_002243, VSP_002244;
CC -!- TISSUE SPECIFICITY: Expressed in various tissues. In hematopoietic
CC tissues, the lymphoid organs express high levels of Meis2. Also
CC expressed in some regions of the brain, such as the putamen.
CC -!- SIMILARITY: Belongs to the TALE/MEIS homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AF179848; AAF81638.1; -
DR EMBL; AF179896; AAF81639.1; -
DR EMBL; AF179897; AAF81640.1; -
DR EMBL; AF179898; AAF81641.1; -
DR EMBL; AF179899; AAF81642.1; -
DR EMBL; BC050431; AAF80431.1; -
DR EMBL; AF017418; AAF80270.1; -
DR HSSP; P40424; I872.
DR TRANSFAC; T04114; -
DR TRANSFAC; T04115; -
DR TRANSFAC; T04116; -
DR TRANSFAC; T04117; -
DR TRANSFAC; T04118; -
DR Genew; HGNC:7001; MEI2.
DR MIM; 601740; -
DR GO; GO:0003704; F:specific RNA polymerase II transcription fa. .; TAS.
DR GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR GO; GO:0000122; P:negative regulation of transcription from P. .; TAS.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SW00389; Hox; 1.
DR PROSITE; PS00027; HOMEBOX 1; FALSE_NEG.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW Homeobox; DNA-binding; Nuclear protein; Alternative splicing.
FT DOMAIN 195 244 SE/THR-RICH
FT DOMAIN 245 273 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 266 273 POLY-ASP.
FT DNA_BIND 276 338 HOMEBOX (TALE-TYPE).
FT VARSPPLIC 346 352 Missing (in isoform Meis2B and isoform Meis2D).
FT VARSPPLIC 301 302 /FTId=VSP 002242.
FT VARSPPLIC 303 477 HP -> VY (in isoform Meis2E).
FT VARSPPLIC 303 477 /FTId=VSP 002243.
FT VARSPPLIC 384 401 Missing (in isoform Meis2E).
FT VARSPPLIC 402 477 LQMPGDYVSGQPGMGMS -> PMSGGMNMGDQWHYM (in isoform Meis2A and isoform Meis2B).
FT VARSPPLIC 402 477 /FTId=VSP 002245.
FT VARSPPLIC 402 477 Missing (in isoform Meis2A and isoform Meis2B).
FT SEQUENCE 477 AA; 51790 MW; 94EBD0801A312B24 CRC64;

Alignment Scores:
Pred. No.: 1.18e-07 Length: 477
Score: 196.50 Matches: 90
Percent Similarity: 33.89% Conservative: 52
Best Local Similarity: 21.48% Mismatches: 152
Query Match: 5.98% Indels: 125
DB: 1 Gaps: 12
US-09-423-575-1 (1-1886) x MEI2_HUMAN (1-477)
QY 460 ATGGGTGGTTCGATTTTCAATTTCTATGCTGCTAACAATAATATCATCTCTCCA 519
Db 14 MetAspGlyValGlyValProAlaSerMetTyrGlyAspProHis-----AlaPro 30
QY 520 CGGTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTTCATCGCTCCTCCTCT 573
Db 31 ArgProIle-----ProProValHis 37
QY 574 -----CATCTCTCCATTCATCTTGGATCATTTTAAGACACTATGATGATTC 624
Db 38 HisLeuAsnHisGlyProProLeuHisAlaThrGlnHisTyrGlyAlaHisAlaProHis 57
QY 625 TCAACACACATG-----TGGGGTTTGAAGCAATAGTAGTTCAGGCATTTTCA 675
Db 58 ProAsnValMetProAlaSerMetGlySerAlaValAsnAlaLeuLeuYsArgAspLys 77

QY 676 GGTGTAGTTGGTCCAAAGTGAACCAATGATGCTACATTCGGTGAAGAAGATTCCCGTTT 735
Db 78 AspAlaIle-----TyrGlyHisProLeuPheProLeu 88
QY 736 CTAATTTCCGAATAAAAGAAACAATGAGCTTTTCATTGAGT----- 774
Db 89 LeuAlaLeuValPheGluLysCysGluLeuAlaThrCysThrProArgGluProGlyVal 108
QY 775 CTTGCATCAGATGTT-----TCTGATGAATGCTCGGAGATAAGTCTTTCTAATAAGTGTGTTACT 828
Db 109 AlaGlyGlyAspValCysSerSerAspSerPheAsnGluAspIleAlaValPheAlaLys 128
QY 829 AGATTAGCTCAGAGCAAGCTTCTTGCAGCAGCAAGACATTTCTAATAAGTGTGTTACT 888
Db 129 GlnValArgAlaGluLysProLeuPheSerSerAsnProGluLeuAspAsnLeuMetIle 148
QY 889 CAAGGTTTCTCTCAACTTATATTGGCTCAAAATACCTTCACTCTGTGTCAAGAATACTA 948
Db 149 GlnAlaIleGlnValLeuArgPheHisLeuLeuGluLeuGluLysValHisGluLeuCys 168
QY 949 TCTCATTTTCGGCGGATCTCGCTC----- 972
Db 169 AspAsnPheCysHisArgTyrIleSerCysLeuLysGlyLysMetProIleAspLeuVal 188
QY 973 -----GATTATTTCATCTCGAGAACCGAGTCAGGAGCTGTAGTTACGCCCTTT 1020
Db 189 IleAspGluArgAspGlySerSerLysSerAspHisGluLeuSerGlySerSerThr 208
QY 1021 ACTTCACCTTTTGAGATATAACTGATTTCTTGTGATGATTAATAACTCGAGGCG 1080
Db 209 AsnLeuAlaAspHisAsnProSerSerTrpArgAspHisAspAlaThrSerThrHis 228
QY 1081 GGTTCGATCTACATTTCAAAGGAGAGCATTAGAAGCAAGAAACCCCTCTCTTGAT 1140
Db 229 SerAlaGlyThrProGlyProSerSerGlyGlyHisAlaSerGlnSer----- 244
QY 1141 CTTCTTCAATGGTGGATGATGATATAGTCATTCGATGATGATGATGATGATGATGAT 1200
Db 245 -----GlyAspAsnSerSerGluGlnGlyAspGlyLeuAspAsnSerVal 259
QY 1201 TCAGCGTTCATGCTGCAACCGAGTTAGATCCACAGTTACACCCCGTTTGCCTCCAA 1260
Db 260 AlaSerProGlyThrGlyAspAspAspProAsp----- 271
QY 1261 ACCGTTTCTTCTTATACAAGAACCTGAGAGAGAGATCTGCAAGAAGATAATCTCTATG 1320
Db 271 ----- 271
QY 1321 GGATCTGTATGGAGAGAGCAAGCAAGACTCAAGAAACCTCTATGTTCCACCAT 1380
Db 272 -----LysAspLysLysArgGln----- 277
QY 1381 TGCCTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTTGGAGACCTCAACGAGGTTG 1440
Db 278 -----LysLysArgGlyLeu 282
QY 1441 ---CCTGAGAAATCTGTTTCGGTTTACGGAATGGATGTTCCAAACTTCTTCCACCT 1497
Db 283 PheProLysValAlaThrAsnIleMetArgAlaThrLeuPheGlnHisLeuThrHisPro 302
QY 1498 TACCCGAAAGATTCCGAGAGAACATCTTCTAGCTATACGAAGTGGCTGACAAAGTCTAG 1557
Db 303 TyrProSerGluGluGlnLysLysGlnLeuAlaGlnAspThrGlyLeuThrIleLeuGln 322
QY 1558 GTATCAAACTGGTTTATAATGCCGGGTTAGCGTATGGAAGCCGATGATGAGAAGAG 1614
Db 323 ValAsnAsnTrpPheIleAsnAlaArgArgIleValGlnProMetIleAspGln 341
RESULT 6
MEI2_MOUSE
ID MEI2_MOUSE STANDARD; PET; 477 AA.
AC P97367; O35676; P97403; P97404;
DT 15-JUL-1998 (Rel. 36, Created)

Db 149 GlnAlaIleGlnValLeuArgPheHisLeuLeuGluLeuGluValHisGluLeuCys 168
QY 949 TCTCATTTCCGCGCATCTCGTC----- 972
Db 169 AspAsnPheCysHisArgTyrIleSerCysLeuLysGlyLysMetProIleAspLeuVal 188
QY 973 -----GATTATTCTATCTCGAGAACCGAGTACGAGCTGTAGTTCAGCCTTT 1020
Db 189 IleAspGluArgAspGlySerSerLysSerAspHisGluLeuSerGlySerSerThr 208
QY 1021 ACTTCAGTTTTCAGATATACTAGTTTCTTGTGATGTTCTTAATCACTCGGAGGG 1080
Db 209 AsnLeuAlaAspHisAsnProSerSerTyrArgAspHisAspAlaThrSerThrHis 228
QY 1081 GGTTCGGATCTACATTTCAAAGGAGGAGGATTAGAAGCAAGAAACCCATCTCTTGGAT 1140
Db 229 SerAlaGlyThrProGlyProSerSerGlyGlyHisAlaSerGlnSer----- 244
QY 1141 CTTCCTCAATGGTGGATGATGATATAGTCATTCGCGTAGATGAGATCATCGGTTATA 1200
Db 245 -----GlyAspAsnSerSerGluGlnGlyAspGlyLeuAspAsnSerVal 259
QY 1201 TCAGCGTTCATCTGCAACCGAGTTAGATCCACAGTTACACACCGGTTTCCCTCCAA 1260
Db 260 AlaSerProGlyThrGlyAspAspAspProAsp----- 271
QY 1261 ACCGTTTCCTCTTATACAGAACCTGAGAGAGAAATCTGCAAGAAGATAATCTCTATG 1320
Db 271 ----- 271
QY 1321 GGATCTGTATTGGAGAGCGCAAGACAGACTCAAGAACCTTATGTTCACAGCAT 1380
Db 272 -----LysAspLysLysArgGln----- 277
QY 1381 TGCCTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACTCAAGAGGTTTG 1440
Db 278 -----LysLysArgGlyIle 282
QY 1441 ----CTGAGAAATCTGTTTCGGTCTACGAAATGGATGTTTCAAAACTTCTTCACCT 1497
Db 283 PheProLysValAlaThrAsnIleMetArgAlaTrpLeuPheGlnHisLeuThrHisPro 302
QY 1498 TACCCGAAGATCGGAGAAACATCTCTAGCTATACGAGTGGCTTGACAAAGATCAG 1557
Db 303 TyrProSerGluGluGlnLysLysGlnLeuAlaGlnAspThrGlyLeuThrIleLeuGln 322
QY 1558 GTATCAAACTGGTTTATAATCGCGGTTAGCTATGGAACCGGATGATAGAAG 1614
Db 323 ValAsnAsnTrpPheIleAsnAlaArgArgGileValGlnPrometileAspGln 341
RESULT 7
ID MEIL_HUMAN STANDARD; PRT; 390 AA.
AC O00470;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homeobox protein Meis1.
GN MEIS1.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9736938; PubMed=9226379;
RA Smith J.E. Jr., Bollekens J.A., Inghirami G., Takeshita K.;
RT "Cloning and mapping of the Meis1 gene, the human homolog of a murine
leukemogenic gene.";
RL Genomics 43:99-103(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;

RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Heltón E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: Expressed at low level in normal
immunohematopoietic tissues, including the fetal liver. Expressed
in a subset of myeloid leukemia cell lines, with the highest
expression seen in those with a megakaryocytic-erythroid phenotype.
CC Also expressed at high levels in the cerebellum.
CC -!- SIMILARITY: Belongs to the TALE/MEIS homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U85707; AAC51642.1; -.
CC EMBL; BC043503; AAH43503.1; -.
CC HSSP; P40424; 1B72.
CC TRANSFAC; T03419; -.
CC Genew; HGNC; 7000; MEIS1.
CC MIM; 601739; -.
CC GO; GO:0003702; F:RNA polymerase II transcription factor acti. .; TAS.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW DNA-binding; Nuclear protein; Homeobox; Proto-oncogene.
FT DOMAIN 194 240 SER/THR-RICH.
FT DOMAIN 242 269 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 262 269 POLY-ASP.
FT DNA_BIND 272 334 HOMEBOX (TALE-TYPE).
SQ SEQUENCE 390 AA; 43016 MW; 3B6A0ACCF0C39121 CRC64;
Alignment Scores:
Pred. No.: 1.36e-07 Length: 390
Score: 195.50 Matches: 97
Percent Similarity: 37.40% Conservative: 50
Best Local Similarity: 24.68% Mismatches: 144
Query Match: 5.95% Indels: 102
Gaps: 14
DB:
US-09-423-575-1 (1-1886) x MEIL_HUMAN (1-390)
QY 574 CATCTCTCTCCACTT-----CATCTTTGGATCATTTAAGACACTATGATTCCTCA 627
Db 41 HisGlyProLeuHisSerHisGlnTyrProHisThrAlaHisThrAsnAlaMetAla 60
QY 628 AACAACTGTGGGGTGTGGTGAAGCAAAATAGTGTTCAGGATTCAGGATTCAGGTGGT 687

Db 61 ProSerMet---GlySerValAsnAspAlaLeuLysArgAspLysAspAlaLeu--- 78
 QY 688 CCAAGTAGCAACCATGATGCTACATTCATGCTGAGAGATTTCCCGTTTCTCTATTTTCCGAAT 747
 Db 79 -----TyrGlyHisProLeuPheProLeuLeuAlaLeuLeu 90
 QY 748 AAAAGAAACAATGACCTTCATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 795
 Db 91 PheGluLysCysGluLeuAlaThrCysThrProArgGluProGlyValAlaGlyGlyAsp 110
 QY 796 GAATGCTCGGATAGCTTT-----TGTGAGCTACAAAGATTGACCTCA 840
 Db 111 ValCysSerSerGluSerPheAsnGluAspIleAlaValPheAlaLysGluIleArgAla 130
 QY 841 GAGCAAGCTTCTTGACGACGACCAAGACATTTCTAATAACGTTGTTTACTCAAGGTTTCTCT 900
 Db 131 GluLysProLeuPheSerSerAsnProGluLeuAspAsnLeuMetIleGlnAlaIleGln 150
 QY 901 CAACTTATATTGGCTCAAAATACCT 957
 Db 151 ValLeuArgPheHisLeuLeuGluLeuGluLysValHisGluLeuCysAspAsnPheCys 170
 QY 958 -----GCCGATACCTCGCTCGATTTATCTATCTCTCGAGAAC 993
 Db 171 HisArgTyrIleSerCysLeuLysGlyLysMetProIleAspLeuValIleAspArg 190
 QY 994 GAGTCAGAGCTGCTAGT---TCAGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1050
 Db 191 GluGlyGlySerLysSerAspSerGluAspIleThrArgSerAlaAsnLeuThr 208
 QY 1051 CTGTGATGCTGATTAATAACTCGAGGCGGTTTCTGATCTACATTTCAAAGGAGACA 1110
 Db 209 ---AspGlnProSerTrpAsn-Arg-----Asp-HisAspAspThrAlaSer 223
 QY 1111 TTAGAAGCAAGAAACCA-----TCTTGGATCTCTCTCAATGGTGGATCAT 1161
 Db 223 hrArgSerGlyGlyThrProGlyProSerSerGlyGlyHisThrSerHisSerGly-Asp 242
 QY 1162 CGATATAGTCATGCTGATGATTCATACCGTTATATACGCTTATCTCTCTCTCTCTCTCTCTCT 1221
 Db 243 AsnSerSerGluGlnGlyAspGlyLeuAspAsnSerValAlaSerProSerThrGlyAsp 262
 QY 1222 GAGTTAGATCCACGATTACACACCCCGTTTGCCTCCAAACCGTTTCTCTCTCTCTCTCTCTCT 1281
 Db 263 AspAspAspProAsp----- 267
 QY 1282 AACCTGAGAGAGAAATCTGCAAGAGAAATATCTCTATGGATCTGTATTGGAGAGAGGC 1341
 Db 267 ----- 267
 QY 1342 AAAGACAAGACTCAAGAAACCTTATGTCTCCACGACATTGCTCTCTCTCTCTCTCTCTCTCTCT 1401
 Db 268 LysAspLys----- 270
 QY 1402 CGAAGAAACATCAGATTTGGACCTCAACAGAGTTTG---CCTGAGAAATCTCTCTCTCTCTCT 1459
 Db 271 ---LysArgHisLys-----LysArgGlyIlePheProLysValAlaThrAsn 285
 QY 1459 GTTCTACGGAATGGATGTTTCAAAACT 1518
 Db 286 IleMetArgAlaAlaTrpLeuPheGlnHisLeuThrHisProTyrProSerGluGluGlnLys 305
 QY 1519 CATCTTCTAGCTATACGAATGGCTGTGACAGAAGTCAGGTATCAACTGTTTATAAT 1578
 Db 306 LysGlnLeuAlaGlnAspThrGlyLeuThrIleLeuGlnValAsnAsnTrpPheIleAsn 325
 QY 1579 GCGCGGTTAGGCTATGGAACCGGATGATGAGAGATGTATGCGGAATGAAACAGAGG 1638
 Db 326 AlaArgArgGlyIleValGlnProMetIleAspGlnSerAsnArgAlaValSerGlnGly 345
 QY 1639 AAGCTCAATCAACAGTCACATTTCAACCAACGGA 1671

Db 346 ThrProTyrAsnProAspGlyGlnProMetGly 356
 RESULT 8
 MEI3 MOUSE
 ID MEI3 MOUSE STANDARD; PRT; 378 AA.
 AC P97368;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Homeobox protein Meis3 (Meis1-related protein 2).
 GN MEIS3 OR MRG2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss Webster;
 RX MEDLINE=97108670; PubMed=8950991;
 RA Nakamura T., Jenkins N.A., Copeland N.G.;
 RT Identification of a new family of Pbx-related homeobox genes.;
 RL Oncogene 13:2235-2242(1996).
 CC -! SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -! TISSUE SPECIFICITY: Expressed at high levels in the brain.
 CC Significant expression also observed in the heart, spleen and lung.
 CC -! DEVELOPMENTAL STAGE: Not expressed until 11 days in embryonic development.
 CC -! SIMILARITY: Belongs to the TALE/MEIS homeobox family.
 CC -! SIMILARITY: Contains 1 homeobox domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U57344; AAC52949.1; --
 DR HSPSP; P40424; I372
 DR TRANSPAC; T03406; --
 DR MGD; MG1108519; Mr92.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SMC0389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX 1; FALSE_NEG.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 KW DNA-binding; Nuclear protein; Homeobox.
 FT DOMAIN 184 232 SER/THR-RICH.
 FT DOMAIN 234 262 ASP/GLU-RICH (ACIDIC).
 FT DNA BIND 265 327 HOMEBOX (TALE-TYPE).
 SQ SEQUENCE 378 AA; 41756 MW; 2175D0F4341E2484 CRC64;
 Alignment Scores:
 Pred. No.: 1,77e-07 Length: 378
 Score: 194.00 Matches: 103
 Percent Similarity: 34.38% Conservative: 50
 Best Local Similarity: 23.15% Mismatches: 137
 Query Match: 5.91% Indels: 155
 DB: 1 Gaps: 15
 US-09-423-575-1 (1-1886) x MEI3_MOUSE (1-378)
 QY 598 GATCATTTAAGACATCATGATGATCTCTCAACACACATGTGGGTTTTGAGCAAAATAGT 657
 Db 6 AspGluLeuArgHisTyrProGlyIleThrGluHisThrThrAlaLeuAlaSerPheSer 25
 QY 658 GAGTTTCAGGCATTTTCAGGTGTAGTTGGTCCCAAGTGAACCAATGATG----- 705
 Db 26 Glu-----AlaAlaProSerValProArgAlaProGlyProTyr 38

```

QY 705 ----- 705
Db 39 ThrProHisArgProProGlnLeuGlnAlaProGlyLeuAspSerLeuLysArg 58
QY 706 -----TCTACATTGGTGAAGAGATTTCCTGCTTCTTAATTTTCGAATAAAGAAC 756
Db 59 GluLysAspAspIleTyrGlyHisProLeuPheProLeuLeuAlaLeuValPheGluLys 78
QY 757 AATGAGCTT-----TCATTGAGTCTTCATCA----- 783
Db 79 CysGluLeuAlaThrCysSerProArgAspGlyAlaSerAlaGlyLeuGlySerProPro 98
QY 784 -----GATGTT-----TCTGATGAATGCTCGAGATAAGTCTTTGTGACGATCAAGA 831
Db 99 GlyGlyAspValCysSerSerAspSerPheAsnGluAspIleAlaPheAlaLysGln 118
QY 832 TTAGCTTCAGACCAAGCTCTTCACAGACAAAGACATTCTTAATTAACCTGTTACTCAA 891
Db 119 IleArgSerGluArgProLeuPheSerSerAsnProGluLeuAspAsnLeuMetValGln 138
QY 892 GGTTCCTCTCAACTATATTGGCTCAAAATACCTTCACCTGTTTCAAGAAATACCTATCT 951
Db 139 AlaIleGlnValLeuArgPheHisLeuLeuGluLeuGluLysValHisAspLeuCysAsp 158
QY 952 CATTCGCGGCATACCTGCTCGATTATTATTCATCTCGAGAACCGAGTCAGAGCTGCTAGT 1011
Db 159 AsnPheCys----- 161
QY 1012 TCAGCCTTACTTCAGCTTTTGAGATATAACTGAGTTCCTGATGGTCAATCTAATAAC 1071
Db 161 ----- 161
QY 1072 TCGAGCGCGGTTTCGATCTACATTTCAAAGGAGAGCATTAGAAGCAAGAAACCCAT 1131
Db 162 -----HisArgTyrIleThrCysLeuLysGlyLys 171
QY 1132 CTC---TTGATATCTTCAATGGTGGATGATCGATATAGTCATTCGCTAGATGAGATT 1188
Db 172 MetProIleAspLeu-----ValIleGluAspArgAspGlySerCysArgGluAspLeu 189
QY 1189 -----CATACGGTTATATACAGGTTTC 1209
Db 190 GluAspTyrThrAlaSerCysProSerLeuProAspGlnAsnThrThr-TripIleArgAs 209
QY 1210 CATGCTGCACCGAGTTAGATCCACAGTTACACCCGGTTTGCCTCCAAACCGTTTCC 1269
Db 209 pHisGluAspSerGlySerValHisLeuGlyThrProGly---ProSer-SerGlyGlyL 228
QY 1270 TTCTTTATACAGAACCTGAGAGAGAATCTGCAAGAGATAATCTCTATGGGATCTGTA 1329
Db 228 euAlaSerGlnSerGlyAspAsnSerSerAspGlnGlyAsp----- 241
QY 1330 TTGGAGAGAGCGCAAGACAGACTCAAGAAACCTCTATGTTCCACCGAGATTGCCTTCTT 1389
Db 242 -----GlyLeuAspThrSerValAlaSerProSerS 252
QY 1390 CACGAGCTGAACGAAGAACCATCATAGATTGAGACCTCAAGAGGT----- 1437
Db 252 erAlaGlyGlu---AspGluAspLeuAspLeuGluArg-ArgArgAsnLysLysArgGly 270
QY 1438 ---TTGCTCGAGAAATCTGTTTCGCTTCTACGAAATTGGATGTTCCAAAACCTCCCTCA 1494
Db 271 IlePheProLysValAlaThrAsnIleMetArgAlaTrpLeuPheGlnHisLeuSerHis 290
QY 1495 CTTTACCGGAAGAGATCGGAGAAACATCTTCTAGCTATACGAGTGGCTGTGACAGAAAT 1554
Db 291 ProTyr-ProSerGluGluGlnLysLysGlnLeuAlaGlnAspThrGlyLeuThrIleLeu 310
QY 1555 CAGGTATCAACCTGGTTTATAATGCGCGGTTAGGCTATGGAAGCGGATGATAGAGAG 1614
Db 311 GlnValAsnAsnTrpPheIleAsnAlaArgArgIleValGlnProMetIleAspGln 330
QY 1615 ATGTATCGCGGAATGAACAGAGGAAGCTCAATACAGAGTCACATTCAACCCACGACCA 1674

```

```

Db 331 -----SerAsnArgThrGlyGlnGlyAlaSerPheAsnProGluGlyGln 345
QY 1675 -----ACTCTCGAATGCCAAATCTGTT 1698
Db 346 ProMetAlaGlyPheThrGluThrGlnProGlnValThrValArgThrProGlySerMet 365
QY 1699 ATGATGACC 1707
Db 366 GlyMetAsn 368
RESULT 9
MEIL_XENLA
ID MEIL_XENLA STANDARD; PRT; 390 AA.
AC P79937; P79938;
DT 15-JUL-1998 (rel. 36, Created)
DT 15-JUL-1998 (rel. 36, Last sequence update)
DE 10-OCT-2003 (rel. 42, Last annotation update)
DE Homeobox protein Meis1 (XMeis1).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC xenopodinae; xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS MEIS1-1 AND MEIS1-2).
RX MEDLINE=97202105; PubMed=9049632;
RA Steelman S., Moskow J.J., Muzynski K., North C., Druck T.,
RA Montgomery J.C., Huebner K., Daar I.O., Buchberg A.M.;
RT "Identification of a conserved family of Meis1-related homeobox
genes.";
RL Genome Res. 7:142-156(1997).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=Meis1-1;
CC IsoId=P79937-1; Sequence=Displayed;
CC Name=Meis1-2;
CC IsoId=P79937-2; Sequence=VSP_002241;
CC -!- SIMILARITY: Belongs to the TALE/MEIS homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U68386; AAB19196.1; -.
CC EMBL; U68387; AAB19197.1; -.
CC HSSP; P41778; 1DU6.
CC TRANSFAC; T03390; -.
CC TRANSFAC; T03391; -.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; homeobox; 1.
CC ProDom; PDC00010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.
CC PROSITE; PS00071; HOMEBOX_2; 1.
CC DNA-binding; Nuclear protein; Homeobox; Alternative splicing.
CC DOMAIN 194 240
CC DOMAIN 242 269
CC DOMAIN 262 267
CC FT DOMAIN 272 334
CC FT DNA BIND 272 334
CC FT VARSPLIC 373 390
CC
CC VARIANT 34 34
CC X -> Q (IN MEIS1-2).

```



```

DR EMBL; AJ417081; CAP01142.1; -
DR EMBL; AB065001; BAB83665.1; -
DR TRANSFAC; T05155; -
DR Genew; HGNC:16714; PKNOX2.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Nuclear protein.
FT DNA_BIND 279 338 HOMEBOX.
FT DOMAIN 8 11 POLY-PRO.
FT DOMAIN 425 439 POLY-GLU.
SQ SEQUENCE 460 AA; 50791 MW; AD2553BC17D9D7B0 CRC64;

Alignment Scores:
Pred. No.: 7,12e-07 Length: 460
Score: 186.50 Matches: 78
Percent Similarity: 41.99% Conservative: 74
Best Local Similarity: 21.55% Mismatches: 125
Query Match: 5.68% Indels: 85
DB: 1 Gaps: 16

US-09-423-575-1 (1-1886) x PNX2_HUMAN (1-460)
QY 811 AGTCTTTGTGTCAGCTACAAGATTAGCTTCAGCAAGCTTCTTGACAGCAAGACATT 870
D 811 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 26 SerLysAlaGlnAlaValHisIleSerAlaProSerAlaAlaSerThrProValPro 45
QY 871 TCTAATAAGCTTGTACTCAAGTTTCTCTCAACTATATATTGGCTCAAAATACCTTAC 930
D 871 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 46 SerAlaProIleAspProGlnAlaGlnLeuGluAlaAspLysArgAlaValTyrArgHis 65
QY 931 TCTGTTCAAGATATCTACTCTCAAGTTTCTCTCAACTATATATTGGCTCAAAATACCTTAC 990
D 931 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 66 ProLeuPheProLeuLeuThr---LeuLeuPheGluLysCysGluGlnAlaThrGlnGly 84
QY 991 ACCGATGACGAGCTGCTAGTTTCAGCTTTTACTTTCAGCTTTTTCAGATATAACTGAGTTT 1050
D 991 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 85 SerGluCys--IleThrSerAlaSerPheAspValAspIleGluAsn-----Phe 100
QY 1051 CTTGATGGTGATTCTTAATACTCGGAGCGGGTTTCGGA-----TCTACA 1095
D 1051 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 101 ValHisGlnGlnGlnGlnGlnHisLysProPheSerAspAspProGluLeuAspAsn 120
QY 1096 TTTCAAAGAGAGACATTAGAAGCAAAAGAAACCCATCTCTTGATCTTCTCAAAATGGTG 1155
D 1096 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 121 LeuMetValLysAlaIleGlnValLeuArgIleHisLeuLeuGluLeuGluLysValAsn 140
QY 1156 GAT-----GATCATATAGTCATTGCGTA----- 1179
D 1156 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 141 GluLeuCysLysAspPheCysAsnArgTyrIleThrCysLeuLysThrLysMethHisSer 160
QY 1180 ---GATGAGATTTCATCGGTTTATATCAGCGTTCCATCGTCAACCGAGTAGATCCACAG 1236
D 1180 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 161 AspAsnLeuLeuArgAsnAspLeuGlyGlyProTyrSerProAsnGlnProSerLeuAsn 180
QY 1237 TTACACACCCGGTTTCCCTCCAA-----ACGTTTCTCTTATACAGAAC 1284
D 1237 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 181 LeuHisSerGlnAspLeuLeuGlnAsnSerProAsnSerMetSerGlyValSerAsnAsn 200
QY 1285 -----CTGAGAGAGAGA 1296
D 1285 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 201 ProGlnGlyIleValValProAlaSerAlaLeuGlnGlnGlnAlaMetThrThr 220
QY 1297 ATCTGCAAGAGATAATCTATGGGATCT-----GTATGGAG 1335
D 1297 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 221 ValAsnSerGlnValValSerGlyGlyAlaLeuTyrGlnProValThrMetValThrSer 240
QY 1336 AGAGGCAACACAAAGCTCAAGAA----- 1359
D 1336 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 241 GlnGlyGlnValValThrGlnAlaIleProGlnGlnGlnAlaIleGlnIleGlnAsnThrGln 260

```

```

QY 1360 -----ACCTCTATGTTCCACGACATTGCTCTTCTTACGACGCTGAACGA 1404
D 1360 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 261 ValAsnLeuAspLeuThrSerLeuLeuAspAsn-----GluAspLysLysSer 276
QY 1405 AAGAACCATCATGATTTCGAGACCTCAACGAGT---TTGCCTGAGAAATCTGTTTCGTT 1461
D 1405 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 277 LysAsn-----LysArgGlyValLeuProLysHisAlaThrAsnIle 290
QY 1462 CTACGGAATTCGATGTTCCAAACTTCTTCCACCTTACCGAAAGATTTCGAGAAACAT 1521
D 1462 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 291 MetArgSerTrpLeuPheGlnHisLeuMethHisProTyrProThrGluAspGluLysArg 310
QY 1522 CTTCTAGCTATACGAAGTGGCTTCACAAGAAGTCAGGTTCAGACTTCAGTTCATTAATGCG 1581
D 1522 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 311 GlnIleAlaAlaGlnThrAsnLeuLeuGlnValAsnAsnTrpPheIleAsnAla 330
QY 1582 CGGTTAGGCTATGGAAGCCGATGATAGAAGAGATGATGCGAA-----ATGAACAAG 1635
D 1582 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 331 ArgArgArgIleLeuGlnProMetLeuAspAlaSerAsnProAspProAlaProLysAla 350
QY 1636 AGGAAGCTCAATACAGTCACATTCACCCCAACGACCACTCTTCAATG---CCAAA 1692
D 1636 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 351 LysLysIleLysSerGlnHis-----ArgProThrGlnArgPheTrpProAsn 366
QY 1693 TCTGTT 1698
D 1693 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 367 SerIle 368

RESULT 11
MEI3_HUMAN
ID MEI3_HUMAN STANDARD; PRT; 382 AA.
AC Q99687; Q9NPW2;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homeobox protein Meis3 (Meis1-related protein 2) (Fragment).
GN MEI3 OR MRG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Blum H., Bauersachs S., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 175-382 FROM N.A.
RX MEDLINE=97202105; PubMed=9049632;
RA Steelman S., Moskow J.J., Muzynski K., North C., Druck T.,
RA Montgomery J.C., Huebner K., Daar I.O., Buchberg A.M.;
RT "Identification of a conserved family of Meis1-related homeobox
genes";
RL Genome Res. 7:142-156(1997).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the TALE/MEIS homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/ebis/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; AL359938; CAB95771.1; -.
CC EMBL; U68385; AB19195.1; -.
CC HSRF; P40424; 1B72.
CC TRANSFAC; T03412; -.
CC Genew; HGNC:7002; MEI3.
CC GO; GO:0005634; C:nucleus; ISS.

```

This SWISS-PRCT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

EMBL; L13663; AAA20882.1; -.
PIR; S42543; S42543.
HSSP; P41778; 1DU6.
TRANSFAC; T04052; -.
InterPro; IPR005539; ELK.
InterPro; IPR001356; Homeobox.
InterPro; IPR005540; KNOX1.
InterPro; IPR005541; KNOX2.
Pfam; PF03789; ELK, 1.
Pfam; PF00046; homeobox; 1.
Pfam; PF03790; KNOX1; 1.
Pfam; PF03791; KNOX2; 1.
ProDom; PD000010; Homeobox; 1.
SMART; SM00389; HOX, 1.
PROSITE; PS00027; HOMEBOX_1; 1.
PROSITE; PS0077; HOMEBOX_2; 1.
Homeobox; DNA-binding; Nuclear protein; Transcription regulation;
Activator.
DOMAIN 102 107 POLY-HIS.
DOMAIN 109 113 POLY-ASN.
DOMAIN 115 122 POLY-SER.
DOMAIN 244 248 ASP/GLU-RICH (ACIDIC).
DOMAIN 259 282 ELK DOMAIN.
DOMAIN 283 345 HOMEBOX (TALE-TYPE).
DNA BIND 379 AA; 42374 MW; 886B020029EDD584 CRC64;
SEQUENCE

```

Alignment Scores:

Alignment Scores:		
Pred. No.:	3.82e-06	379
Score:	177.00	Length:
Percent Similarity:	35.31%	Matches:
Best Local Similarity:	20.18%	Conservative:
Query Match:	5.39%	Mismatches:
DB:	1	Indels:
		Gaps:

US-09-423-575-1 (1-1886) x HMB1 SOYBN (1-379)

QY	298	AACCCAAATGGACAACAACAACAACAACACTTTT	TAGTTCTCTGGATAATGTCAGACT	357
Dd	42	HisProIleAsnProSerAsnAsnAsn-----	Val 52	
QY	358	ACCACAAATCCTCTTCTCATGGATTATATACCTTCAAGAGAAGATCAACTTCATCTCA	417	
Dd	53	AsnThrAsnCysLeu-----PheIleProAsnCysSerAsnSerThrGlyThrPro	69	
QY	418	ACAATGCTTCCAATGAATACCATCAGATCAGATCCTCTCACAAATGGGTGGCTTTGATATT	477	
Dd	70	SerIleMet-----	Leu 73	
QY	478	TTCAATTTCTATCTCGACTAACAAATACATTATCACTTCTCCACGGTCTATCGATGTTCAA	537	
Dd	74	HisAsnAsnHisAsnAsnAsnLys-----	ThrAspAspAsp 85	
QY	538	GATAACCGGAAT-----GTTGAGTTTCATGGCTCCTCCTCTCATCTCTCT	582	
Dd	86	AspAsnAsnAsnAsnThrGlyLeuGlyTyrTyrPheMetGlusEr	100	
QY	583	CCACTTCATCCTTTGGATCATTTAAGACACACTATGATGATTCCTCAACAACATGTGGGT	642	
Dd	101	-----AspHisHisHisHisHisGlyAsnAsnAsnAsnAsn-----	Gly 114	
QY	643	TTTAGACAAAATAGTGAGTTTCAGGCATTTTCAGGTGTAGTTGGTCCAGTGAACCAATG	702	
Dd	115	SerSerSerSerSerSerSerSerVallysAlalysile-----	128	

QY	703	ATGTCATACATTCCGCTGAAGAAGATTTCCTCCGTTTCTTAATTTTCGAATAAAGAAACAATGAG	762
Db	129	-----MetAlaHisProHisTyrHisArgLeuLeuAlaalaTyrValAsnCysGln	145
QY	763	CTTTCATTGAGTCTTGCATCAGATGTTTCTGATGAATGCTCGGAGATAAGCTTTTGTGCA	822
Db	146	---LysValGlyAlaProProGluValValAlaArgLeuGluGluAlaCysAlaSerAla	164
QY	823	GCTACAAGATTAGCCTCAGACCAAGCT-----TCATTGCAGCAGCAAGAATTCCT	873
Db	165	AlaThrMetAlaGlyGlyAspAlaAlaAlaGlySerSerCysIleGlyGluAspProAla	184
QY	874	AATAACGTTGTTACTCAAGGTTTCTCTCAACTTATATTTGGCTCAAAATACCTTCACCTCT	933
Db	185	LeuAspGlnPheMetGluAlaTyrCysGluMetLeu-----ThrLysTyrGluGlnGlu	202
QY	934	GTTCAAGAAATACATCTATCTATTTTCGGCGCATACTCGCTCGATTATTTCATCTCGAGGAACC	993
Db	203	LeuSerLysProLeuLys-----	208
QY	994	GAGTCAGGAGCTGCTAGTTCAGCTTTCATCTACGTTTTCAGCAATATACTAGTTTCTT	1053
Db	209	GluAlaMetLeuPheLeuGlnArgIleGluCysGlnPheLysAsnLeuThr-----IleSer	227
QY	1054	GATGGTGATCTTAATAACTCGGAGCGGGTTCGGATCTACTATTCAAAGGAGAGCATTA	1113
Db	228	SerSerAspPheAlaSerAsnGluGlyGly-----	237
QY	1114	GAAGCAAGAAACCCCATCTCTTTGGATCTTCTTCAATGGTGGATGATCGATATAGTCAT	1173
Db	238	-----	242
QY	1174	TGCGTAGATCAGATT-----CATACGGTTTATATCAGCGTTCCATGCTCGCAACCGAGTTA	1227
Db	243	SerGluGluAspValAspLeuHisAsnMetIle-----	253
QY	1228	GATCCACAGTTACACACCGGGTTTGGCCCTCCAAACCGTTTCTCTTATACAAGAACCTG	1287
Db	254	AspProGlnAlaGluAspArg-----	262
QY	1288	AGAGAGAGATCTGCACAGAGATAAATCTCT--ATGGGATCTGTATTGGAGAGAGGCAAA	1344
Db	263	LysGlyGlnLeuLeuArgLysTyrSerGlyTyrLeuGlySer-----	276
QY	1345	GACAAGACTCAAGAAACCTCTATGTTTCCACCAGCATTCGCTTCTTTCAGCAGCTGAAACGA	1404
Db	277	-----LeuLysGlnGluPheMetLys	283
QY	1405	AAGAACCATCAGATTGGAGACCTCAACGAGGTTTCGCTGAGAAATCTGTTTCGTTCTTA	1464
Db	284	Lys-----ArgLysLysGlyLysLeuProLysGluAlaArgGlnGlnLeu	298
QY	1465	CGGAATTGGATGTTCCAAACCTTCCTTCCACCTTACCAGCAAGATTTCGAGAGAACATCTT	1524
Db	299	LeuGluTrpTrpAsnArgHisTyrLysTyrProTyrProSerGluSerGlnLysLeuAla	318
QY	1525	CTAGCTATACGAAGTGGCTTGACAGAAGTCAGGTATCAAACTGGTTTATAATATGCGCGG	1584
Db	319	LeuAlaGluSerThrGlyLeuAspGlnLysGlnIleAsnAsnTrpPheIleAsnGlnArg	338
QY	1585	GTTAGGCTATGGAAGCGCATGATAGAAGATCTATCGGAATGAAC	1632
Db	339	LysArgHisTrpLysProSerGluAspMetGlnPheValValMetAsp	354

RESULT 13

HLK6 LYCES

ID_HLK6 LYCES STANDARD; PRT; 355 AA.

AC O22289;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Homeobox protein knotted-1 like LET6.

DN LET6.

RESULT 13

RESULT 13	STANDARD	PRT
HKL6_LYCES		355 AA
ID_HKL6_LYCES		

ID	HRL6 LYCES	STANDARD;	PRT;	355 AA.
AC	022299;			
DT	15 JUL 1968	15-1	20	022299

DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update
DE Homeobox protein knotted-1 like LET6.

DE HOMEOBOX PROTEIN ANTICOD-1 LIKE ZERO;
GN LET6.

29. After

Db 291 GlnlyseuAlaLeuAlaGluSerThrGlyLeuAspGlnHisGlnleAsnAsnTrpPhe 310
QY 1573 ATAAATCGCGGGTATAGCTATGAGCGCATGATAGAGAGATGATCGGAAATGAAC 1632
Db 311 IleAsnGlnArgIleHisTrpLysProSerGluAspMetGlnPheValMetAsp 330
RESULT 14
HKL3 ARATH
ID HKL3 ARATH STANDARD; PRT; 431 AA.
AC P48000;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Homeobox protein knotted-1 like 3 (KNAT3).
GN KNAT3 OR AT5G25220 OR F2AJ6.18
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97304677; PubMed=9161040;
RA SriKawa K.A.; Martinez-Iaborda A.; Kim H.S.; Zambryski P.C.;
RA "Localization of expression of KNAT3, a class 2 knotted1-like gene.";
RL Plant J. 11:853-861(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S.; Kaneko T.; Nakamura Y.; Kotani H.; Kato T.; Asamizu E.;
RA Miyajima N.; Sasamoto S.; Kimura T.; Hosouchi T.; Kawashima K.;
RA Kohara M.; Matsumoto M.; Matsuno A.; Muraki A.; Nakayama S.;
RA Nakazaki N.; Naruo K.; Okumura S.; Shinpo S.; Takeuchi C.; Wada T.;
RA Watanabe A.; Yamada M.; Yasuda M.; Sato S.; de la Bastide M.;
RA Huang E.; Spiegel L.; Gnoj L.; O'Shaughnessy A.; Preston R.;
RA Habermann K.; Murray J.; Johnson D.; Rohlfing T.; Neilson J.;
RA Stoneking T.; Pupin K.; Spieth J.; Sekhon M.; Armstrong J.; Becker M.;
RA Belter E.; Cordum H.; Cordes M.; Courtney L.; Courtney W.; Dante M.;
RA Du H.; Edwards J.; Fryman J.; Haakensen B.; Lamar E.; Latreille P.;
RA Leonard S.; Meyer R.; Mulvaney E.; Ozersky P.; Riley A.; Strommatt C.;
RA Wagner-McPherson C.; Wollam A.; Yeakum M.; Bell M.; Dedhia N.;
RA Parnell L.; Shah R.; Rodriguez M.; Hoon See L.; Vil D.; Baker J.;
RA Kirchoff K.; Roth K.; King L.; Bahret A.; Miller B.; Marita M.A.;
RA Martienssen G.; McComb W.R.; Wilson R.K.; Murphy G.; Bancroft I.;
RA Volckaert G.; Wambutt R.; Duesterhoeft A.; Stiekema W.; Pohl T.;
RA Entian K.-D.; Terryn N.; Hartley N.; Bent E.; Johnson S.;
RA Langham S.-A.; McCullagh B.; Robben J.; Grynoprez B.; Zimmermann W.;
RA Ransberger U.; Wedler H.; Balke K.; Wedler E.; Peters S.;
RA van Staveren M.; Dirkse W.; Mooljman P.; Klein Lankhorst R.;
RA Weitzensger T.; Bothe G.; Rose M.; Hauf J.; Bernerstorfer S.; Hempel S.;
RA Feldpausch M.; Lamberth S.; Villarreal R.; Gielens J.; Ardiles W.;
RA Bents O.; Lemcke K.; Kolesov G.; Mayer K.; Rudd S.; Schoof H.;
RA Schueller C.; Zaccaria P.; Mewes H.-W.; Bevan M.; Franz P.;
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT thaliana".
RL Nature 408:823-826(2000).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the TAIR/KNOX homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X92392; CAA63130.1; --
CC EMBL; AC008259; AAC98441.1; --
CC TRANSFAC; T04044; --
CC InterPro; IPR005539; ELK.

DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR005540; KNOX1.
DR InterPro; IPR005541; KNOX2.
DR Pfam; PF03789; ELK; 1.
DR Pfam; PF03790; KNOX1; 1.
DR Pfam; PF03791; KNOX2; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW Homeobox; DNA-binding; Nuclear protein.
FT DOMAIN 23 33 POLY-PRO.
FT DOMAIN 34 37 POLY-GLN.
FT DOMAIN 58 61 POLY-ASN.
FT DOMAIN 84 90 POLY-ALA.
FT DOMAIN 104 108 POLY-ASN.
FT DOMAIN 320 343 ELK DOMAIN.
FT DNA BIND 344 406 HOMEBOX (TAIR-TYPE).
SQ SEQUENCE 431 AA; 4760 MW; 5222B67AB54B9673 CRC64;
Alignment Scores:
Pred. No.: 5.13e-06 Length: 431
Score: 175.50 Matches: 93
Percent Similarity: 34.80% Conservative: 57
Best Local Similarity: 21.58% Mismatches: 206
Query Match: 5.34% Indels: 75
DB: 1 Gaps: 13
US-09-423-575-1 (1-1886) x HKL3 ARATH (1-431)
QY 523 TCTATCATGTTTCAAGATAACCGCAATGTTGAGTTCATGGCTCCTCCTCATCCTCTCT 582
Db 9 SerGlnAspLeuSerPheAsnHisPheThrAspGlnHisGlnProProProGlnPro 28
QY 583 CCATCTCATCCTTTGGATCATTTAAGACATAT----- 615
Db 29 ProProProProGlnGlnGlnHisPheGlnGlnAlaProProProAsnTrpLeu 48
QY 616 -----GATGATTCTCTCAACCAACATGCTGGGTTTTTGAA---GCA 651
Db 49 AsnThrAlaLeuLeuArgSerSerAspAsnAsnAsnAsnPheLeuAsnLeuHisThrAla 68
QY 652 AATAGTGAAGTTTCAAGGATTTTCAAGTGTAGTTGGTCCAAAGTGAACCAATGATCTTACA 711
Db 69 ThrAlaAsnThrThrThrAlaSerSerSerSerSerSerSerSerSerSerAlaAlaAla 88
QY 712 TTCGGTGAAGAA-----GATTCCCGTTCTTAATTCGAATAAAGAAAGCAAT 759
Db 89 AlaAlaAsnGlnTrpLeuSerArgSerSerSerPheLeuGlnArgAsnAsnAsnAsn 108
QY 760 GAGCTTTTCATGAGTCTTGCATCAGATGTTTCTGATGAATGCTCGAGATAAGTCTTGT 819
Db 109 -----AlaSerIleValGlyAspGlyIleAspValThrGlyGly 122
QY 820 GCAGCTACAGATTAGCTCAGACAGCTTCTTGCAGCAGCAAGACATTTCTAATAAC 879
Db 123 AlaAspThrMetIleGlnGlyGluMetLysThrGlyGlyGlyGlyGlyGlyGlyGly 142
QY 880 GTTGTACTCAAGGTTTCTCTCAACTATATTTGGCTCAAAATACCTCTACTCTGTTCAA 939
Db 143 GlyGlyAlaThrAlaAlaAspGlyValValSerTrpGlnAsnAlaArgHisLysAla--- 161
QY 940 GAAATATCTATCTAT-----TTC 957
Db 162 GluIleLeuSerHisProLeuTrpGluGlnLeuLeuSerAlaHisValAlaCysLeuArg 181
QY 958 GCCCATACTCGCTCGATTATTTCATCTCGAGGAGACCGAGTCAGAGCTGCTAGTTCAGCC 1017
Db 182 IleAlaThrProValAspGlnLeuProArgIleAspAlaGlnLeuAsnSerGlnHis 201
QY 1018 TTTACTTCACGTTTGTGAGATAATACTAGCTTTCTTGTGATGGTGTCTTAATAACTCGGAG 1077
Db 202 ValValAlaLysTyrSerAlaLeu-----GlyAla 211

Db	26	SerSerProLeuSerAlaValIleSerProProGlnGlnGlnGlnHisGlnGln	45
Qy	721	GAAGATTTCGGTTTCTTAATTTCGAATAAAGAAACAATGAGCTTTTCATTGAGCTTGCA	780
Db	46	GlnSerAlaGlyTyrLeu-----AlaHisSerProLeuSerLeuAenThrAla	61
Qy	781	---TCAGATGTTCT-----GATCAATGCTCGAGATAACTCTTTGTGCAGCT	825
Db	62	ProProGlyValSerHisGlyGlySerGlyCysSerAsnProValLeuGlnLeuAla	81
Qy	826	ACAAGATTAGCCTCAGAGCAAGCTTCCTGAGACGACAAAGACATTTCTTAATAACAGTTGTT	885
Db	82	AsnGlySerLeuLeuGluAlaCysAlaAlaAlaLysGluProSerSer-----	98
Qy	886	ACTCAAGTTTCTCTCAACTATATTTGGCTCAAAATACCTTCACTCTGTTCAA-----	939
Db	99	-----SerSerTyrAlaAlaAspValGluAlaIle	108
Qy	940	-----GAAATACTATCTCAT-----TTGCGCGCATACTCGCTCGAT	975
Db	109	LysAlaLysIleIleSerHisProHisTyrSerSerLeuLeuAlaIaTyr---LeuAsp	127
Qy	976	TATTCACTCTCGAGAACCGAGTCAGAGCTGCTAGTTTACGCGCTTACTTCACGTTTGTAG	1035
Db	128	CysGlnLysValGly-----AlaProGluValSerAlaAspLeuThr	142
Qy	1036	AATATAACTGAGTTTCTTGATGGTGATTCTAATAACTCGGAGCGGGTTTCGATCTACA	1095
Db	143	AlaValAlaGlnAspLeuLeuArgGlnArgThrAlaLeuGlyGlyLeuGlyThrAla	162
Qy	1096	TTTCAAAGAGAGCATTAGAAAGCAAGAAACCCATCTCTTGGATCTTCTCAATGGTG	1155
Db	163	ThrGlu-----	164
Qy	1156	GATGATCGATATAGTCATTGGGTAGATGAGATTTCATACGGTTATATACAGGTTCCATGCT	1215
Db	165	-----ProGluLeuAspGlnPheMetGluAlaTyrHisGlu	176
Qy	1216	GCAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCCTCCAAACCGTTTCTCTCTTA	1275
Db	177	MetLeuValLysTyrArgGluGluLeuThrArgProLeuGlnGluAlaMetGluPheLeu	196
Qy	1276	TACAAG-----AACCCTGAGAGAGAAATCTCCACAGACAGATA	1311
Db	197	ArgArgValGluThrGlnLeuAsnSerLeuSerIleSerGlyArgSerLeuArgAsnIle	216
Qy	1312	ATCTCTATGGATCTGTATTGGAGAGAGGCAAGACAAAGACTCAAGAAACCTCTATG---	1368
Db	217	LeuSerThrGlySerSerGluLysGlnGlySerGlyGlyGluThrGluLeuPro	236
Qy	1369	---TTCCACCAGCATTCCTTCTTCACGACTGAAACGAAAGAACCATCAGATTCTGGAG	1425
Db	237	GluIleAspAlaHisGlyValaspGlnGluLeuLys-----HisHisLeuLeuLys	253
Qy	1426	CCTCAACGAGGT-----	1437
Db	254	LysTyrSerGlyTyrLeuSerSerLeuLysGlnGluLeuSerLysLysLysLysGly	273
Qy	1438	---TTGCCTCAGAAATCTGTTTCGGTTCTACGGAATTTGATGTTCACAAACTTCCTTCAC	1494
Db	274	LysLeuProLysGluAlaAspGlnGlnLeuLeuSerTyrTrpGluMetHisTyrLysTrp	293
Qy	1495	CCITTACCGGAAGATTCCGGAACAATCTTCTAGCTATACGAAGTGGCTTCACAAAGAT	1554
Db	294	ProTyrProSerGluSerGlnLysValAlaLeuAlaGluSerThrGlyLeuAspLeuLys	313
Qy	1555	CAGGTATCAAACTGGTTTATAAATGCGCGGTTAGGCTATGGAAGCCGATGATAGAAGAG	1614
Db	314	GlnIleAsnAsnTrpPheIleAsnGlnArgHisTrpLysProThrAspGluMet	333
Qy	1615	ATGATATCGCGAATGAAC	1632
Db	334	GlnPheValMetMetAsp	339

mis Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 2, 2004, 11:35:18 ; Search time 118 Seconds
(without alignments)

10085.887 Million cell updates/sec

Title: US-09-423-575-1

Perfect score: 3284

Sequence: 1 attagttataaatgttgc.....ttaaaaaaaaaaaaaaaaaa 1886

Scoring table:

BLOSUM62 Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame*_n2p.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US09423575/runat_01092004_161106_18570/app_query.fasta_1.2055
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09423575.@CGN_1_162.@runat_01092004_161106_18570 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_rvirus: *
16: sp_bacteriopl: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2453	74.7	469	10 Q7X9S7	Q7X9S7 arabidopsis

2	736.5	22.4	323	10 Q7Y0Z6	Q7Y0Z6 lycopersico
3	487	14.8	699	10 Q7Y0Z7	Q7Y0Z7 lycopersico
4	477	14.5	688	10 Q8LLE4	Q8LLE4 solanum tub
5	472	14.4	680	10 Q94KL4	Q94KL4 arabidopsis
6	472	14.4	680	10 Q9SJ56	Q9SJ56 arabidopsis
7	465.5	14.2	535	10 Q8LLE3	Q8LLE3 solanum tub
8	457.5	13.9	567	10 Q8LLE9	Q8LLE9 solanum tub
9	450.5	13.7	393	10 Q7Y0Z8	Q7Y0Z8 lycopersico
10	450.5	13.7	532	10 Q8LLE1	Q8LLE1 solanum tub
11	444	13.5	524	10 Q9FWS9	Q9FWS9 arabidopsis
12	444	13.5	524	10 Q8W412	Q8W412 arabidopsis
13	439	13.4	642	10 Q9AYD9	Q9AYD9 oryza sativ
14	436	13.3	523	10 Q7Y0Z9	Q7Y0Z9 lycopersico
15	427.5	13.0	759	10 Q94KS5	Q94KS5 hordeum vul
16	427	13.0	645	10 Q8LLE8	Q8LLE8 solanum tub
17	426.5	13.0	884	10 Q8S5U8	Q8S5U8 oryza sativ
18	426	13.0	809	10 Q9M7S0	Q9M7S0 malus domes
19	425.5	13.0	538	10 Q9FXG8	Q9FXG8 arabidopsis
20	424	12.9	739	10 Q9SW80	Q9SW80 arabidopsis
21	422.5	12.9	620	10 Q8LLE0	Q8LLE0 solanum tub
22	421.5	12.8	638	10 Q23196	Q23196 arabidopsis
23	419	12.8	482	10 Q9SIW1	Q9SIW1 arabidopsis
24	419	12.8	482	10 Q932M1	Q932M1 arabidopsis
25	418	12.7	611	10 Q38997	Q38997 arabidopsis
26	417.5	12.7	532	10 Q65685	Q65685 arabidopsis
27	417	12.7	513	10 Q93Y33	Q93Y33 arabidopsis
28	414.5	12.6	431	10 Q8S897	Q8S897 arabidopsis
29	408	12.4	627	10 Q64826	Q64826 arabidopsis
30	407	12.4	627	10 Q94KL5	Q94KL5 arabidopsis
31	404	12.3	567	10 Q8LLE2	Q8LLE2 solanum tub
32	404	12.3	586	10 Q8LN25	Q8LN25 oryza sativ
33	403.5	12.3	900	10 Q7XA57	Q7XA57 gnetum gnet
34	401	12.2	584	10 Q9SJJ3	Q9SJJ3 arabidopsis
35	380	11.6	575	10 Q9LZW8	Q9LZW8 arabidopsis
36	379.5	11.6	575	10 Q8H147	Q8H147 arabidopsis
37	379	11.5	575	10 Q8LGG2	Q8LGG2 arabidopsis
38	376.5	11.5	611	10 Q94KS4	Q94KS4 hordeum vul
39	358	10.9	345	10 Q9AYD8	Q9AYD8 oryza sativ
40	348.5	10.6	612	10 Q941S9	Q941S9 oryza sativ
41	348.5	10.6	612	10 Q8RUM5	Q8RUM5 oryza sativ
42	341	10.4	357	10 Q8H693	Q8H693 oryza sativ
43	341	10.4	357	10 Q8H695	Q8H695 oryza sativ
44	338.5	10.3	359	10 Q8RXA9	Q8RXA9 zea mays (m
45	336.5	10.2	92	10 Q42302	Q42302 arabidopsis

ALIGNMENTS

RESULT 1

Q7X9S7	PRELIMINARY;	PRT;	469 AA.
ID Q7X9S7			
AC Q7X9S7			
DT 01-OCT-2003 (TRENBLrel. 25, Created)			
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)			
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)			
DE H1 gene protein.			
GN H1 GENE.			
OS Arabidopsis thaliana (Mouse-ear cress).			
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;			
OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.			
ON NCBI_taxID=3702;			
RX STRAIN=CV, Colombia; TISSUE=Rosette leaves;			
RA MEDLINE=95210927; PubMed=7696878;			
RT "The Homeobox gene AtH1 of Arabidopsis is depressed in the			
RL Plant Cell 7:117-129(1995).			
DR EMBL; X80127; CAC51426.1; -			
SQ SEQUENCE 469 AA; 53477 MW; 3DC7C6A976819574 CRC64;			

Alignment Scores:

Pred. No.: 9-9e-208 Length: 469
 Score: 2453.00 Matches: 469
 Percent Similarity: 99.15% Conservative: 0
 Best Local Similarity: 99.15% Mismatches: 0
 Query Match: 74.70% Indels: 4
 DB: 10 Gaps: 1

US-09-423-575-1 (1-1886) x Q7X9S7 (1-469)

QY	304	ATGCAACAAACAAACAAACAAACATCTTCTCTGATTAATGTCATGACTAACAA	363
Db	1	MetAspAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnValMetThrAsnGln	20
QY	364	AATCCTCTTCTCATGGATTTTATACCTTCAAGAGAGATTCAACTTCATTCACAAATG	423
Db	21	AsnProLeuLeuMetAspPheileProSerArgGluAsp-----SerThrMet	36
QY	424	CTTCCATGAATACCATCATGATCAGATCCTCTACAAATGGGTGGCTTTCATATTTCAAT	483
Db	37	LeuProTrpAsnThrIleArgSerAspProLeuGlnMetGlyGlyPheAspIlePheAsn	56
QY	484	TCTATCTGACTAAACAAATACTTATCATCTCTCCACGGTCTATCGATGTTCAAGATAAC	543
Db	57	SerMetLeuThrAsnLysTyrLeuSerSerSerProArgSerIleAspValGlnAspAsn	76
QY	544	CGCAATGTTGAGTTCATGGCT	603
Db	77	ArgAsnValGluPheMetAlaProProHisProProLeuHisProLeuAspHis	96
QY	604	TTAAGACACTATGATGATTCCTCAACACACATGTGGGGTTTTGAAGCAAAATAGTGAGTTT	663
Db	97	LeuArgHisTyrAspAspSerSerAsnAsnMetTrpGlyPheGluAlaAsnSerGluPhe	116
QY	664	CAGGCATTTTCAGGTGATCTGGTCCAACTGACCAATGATCTACATTCGGTGAAGAA	723
Db	117	GlnAlaPheSerGlyValValGlyProSerGluProMetMetSerThrPheGlyGluGlu	136
QY	724	GATTTCCCGTTCTTAATTTTGAATTAAGAAACATGAGCTTTTCATGAGCTTCGATCA	783
Db	137	AspPheProPheLeuIleSerAsnLysArgAsnAsnGluLeuSerLeuSerLeuAlaSer	156
QY	784	GATGTTTCTGATGATGTCGGAGATAAGTCTTTTGTGCGAGCTACAGATTAGCCTCAGAG	843
Db	157	AspValSerAspGluCysSerGluIleSerLeuCysAlaAlaThrArgLeuAlaSerGlu	176
QY	844	CAAGCTTCTTGACGACGACGACAAACATTTCTATACGTTGTACTCAAGGTTCTCTCAA	903
Db	177	GlnAlaSerCysSerSerLysAspIleSerAsnAsnValValThrGlnGlyPheSerGln	196
QY	904	CTTATATTTGGCTCAAAATACCTTCACTCTGTTCAGAAATACTATCTCATTTCCGCCCA	963
Db	197	LeuIlePheGlySerLysTyrLeuHisSerValGlnGluLeuSerHisPheAlaAla	216
QY	964	TACTGCTCGATATTCATCTCAGAGAACCGAGTACGAGCTGCTAGTTTCAGCCTTTACT	1023
Db	217	TyrSerLeuAspTyrSerSerArgGlyThrGluSerGlyAlaAlaSerSerAlaPheThr	236
QY	1024	TCACGTTTTTGAATATTAATCTCAGTTCTTGTGATGTTCTTAATAACTCGGAGCGGGT	1083
Db	237	SerArgPheGluAsnIleThrGluPheLeuAspGlyAspSerAsnAsnSerGluAlaGly	256
QY	1084	TTCCGATCTACATTTCAAAGGAGACATTTAGAGCAAAAGAAACCCATCTCTTGATCTT	1143
Db	257	PheGlySerThrPheGlnArgAlaLeuGluAlaLysLysThrHisLeuLeuAspLeu	276
QY	1144	CTTCAATGGTGCATCATATAGTATGTCGATGATGATGATGATGATGATGATGATGATGAT	1203
Db	277	LeuGlnMetValAspAspArgTyrSerHisCysValAspGluIleHisThrValIleSer	296
QY	1204	GGTTTCCATGTCGCAACCGAGTTAGATCCACAGTTTACACCCCGGTTTCCCTCCCAACC	1263
Db	297	AlaPheHisAlaAlaThrGluLeuAspProGlnLeuHisThrArgPheAlaLeuGlnThr	316

QY	1264	GTTTCCTTCTTATACAAAGAACCTGAGAGAGAGAAATCTGCAAGAAGATAATCTCTATGGGA	1323
Db	317	ValSerPheLeuTyrLysAsnLeuArgGluArgIleCysLysLysIleSerMetGly	336
QY	1324	TCTGATTTGGAGAGAGCAAGACACTCAAGAAACCTCTATGTTCCACCAAGCATTC	1383
Db	337	SerValLeuGluArgGlyLysAspLysThrGlnGluThrSerMetPheHisGlnHisCys	356
QY	1384	CTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACCTCAACGAGGTTTGCCT	1443
Db	357	LeuLeuGlnGlnLeuLysArgLysAsnHisGlnIleTrpArgProGlnArgGlyLeuPro	376
QY	1444	GAGAAATCTTTTCGGTTCTACGAAATGATGTTCCAAAACCTTCCTCCACCTTACCG	1503
Db	377	GluLysSerValSerValLeuArgAsnTrpMetPheGlnAsnPheLeuHisProTyrPro	396
QY	1504	AAAGATTCGGAGAAACATCTTCTAGTATACGAGTGGCTTGACAAAGAGTCAAGTATCA	1563
Db	397	LysAspSerGluLysHisLeuAlaIleArgSerGlyLeuThrArgSerGlnValSer	416
QY	1564	AACTGCTTTTATAAATCGCGGTTAGGCTATGGAACCGATGATAGAAGATGATGCG	1623
Db	417	AsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluGluMetTyrAla	436
QY	1624	GAATGACAAAGAGGAGCTCAATACAGTCACATTCACATTCACCCACGACCACTCTTCA	1683
Db	437	GluMetAsnLysArgLysLeuAsnAsnSerHisIleGlnProAsnGlyProThrLeuArg	456
QY	1684	ATGCCAAATCTCTTATGATGAGCCAAAGCAATGCATAAA	1722
Db	457	MetProLysSerValMetMetSerGlnAlaMetHisLys	469

RESULT 2

Q7Y0Z6 PRELIMINARY; PRT; 323 AA.
 AC Q7Y0Z6;
 DT 01-OCT-2003 (T-EMBLrel. 25, Created)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Bell-like homeodomain protein 4 (fragment).
 GN Bt4.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ron N., Parnis A., Lifschitz E.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR ENBL; AF375967; AAP47026.1; -;
 KW DNA-binding; Homeobox; Nuclear protein.
 FT NON TER 1 1
 SQ SEQUENCE 323 AA; 37216 MW; DED2C654B526D6F1 CRC64;

Alignment Scores:

Pred. No.: 3-37e-56 Length: 323
 Score: 736.50 Matches: 166
 Percent Similarity: 65.16% Conservative: 36
 Best Local Similarity: 53.55% Mismatches: 72
 Query Match: 22.43% Indels: 37
 DB: 10 Gaps: 7

US-09-423-575-1 (1-1886) x Q7Y0Z6 (1-323)

QY	793	GATGAATGTCGAGAGATAAGTCTTCTGTCAGCTACA-----AGA	831
Db	2	AspGlnCysSerAspIleSerCysSerGlyValThrAsnHisAlaPheProGlnArg	21
QY	832	TTAGCTTCAGAGCAAGCTTCTTCGACGACCAAGACATTTCT-----AATAACGTT	882
Db	22	PheAspSerGluLeuThrSerCysAsnSerArgAsnLeuSerLeuSerPheGlySerTyr	41

883 GTTACTCAAGGTTCTCTCAACTTATATTTGGTCAAAATACCTTCACTCTCTTCAAGAA 942
Dd
42 LysProValTyrLeuSerGlnPheLeuThrGlySerArgTyrLeuArgValMetGlnGlu 61
943 ATACTATCTCATTTCCGCCATACCTCGCTCGAT-----TATTCACT 984
Dd
62 IleLeuSerGluIleAlaGlnLeuSerLeuGlnAsnHisAsnLeuValGlyTyrArgGly 81
985 CGAGGAACCGAGTCAGAGCTGTAGTTTCAGCGCTTTTACCTCAGCTTTTGGAGATTAAT 1044
Dd
82 AsnGlyThrGluAsnGlyAla-----AsnThrSer 91
1045 GAGTTTCTGATGATCTTAATACTCGAGGGGGTTTCGATCTACATTTCAAAGG 1104
Dd
92 PheAlaLeuAsnSerAspAlaGlyArgGlyTyrAlaAlaMetSerSerAspSerPro 111
1105 AGAGCATTAA-----GAAGCAAGAAACCCATCTCTTGGATCTTCTCAATGGTG 1155
Dd
112 AspGlyLeuMetGlyCysGluAlaLysLysAsnLeuValAlaLeuGlnVal 131
1156 GATGATCGATATAGTCATTGCGTAGATGATTCATACGTTTATATACGCGTTCCATGCT 1215
Dd
132 AspAspGlnTyrAsnGlnCysLeuAspGluIleHisMetValIleSerAlaPheHisAla 151
1216 GCACCGAGTTAGATCCAGATAGACACCGGTTTGGCTTCCAAACCGTTTCTCTTA 1275
Dd
152 ValThrGluLeuAspSerIleHisAlaArgPheAlaLeuGlnThrIleSerSerLeu 171
1276 TACAAGAACCTGAGAGA-GAGAACTCTGCAAGAAAGATAATCTCTATGGGATCTGTATGGA 1334
Dd
172 TyrLysAsnLeuArgGlyGluAsnLysGlnSerHisSerArgAsnGlyArgThrPheGln 191
1335 GAGAGG---CAAAGCAAGACTCAAGAAACCTCTATGTTCCACGAGCTTGCCTT----- 1386
Dd
192 GlnArgMetArgLysGlySerArgGluVal-IleLeuLysHisHisSerPheLysE 211
1387 -----CTTACACGCTGAACGAAGAACCATCAGATTTCGAGACCTCAACGAGG 1436
Dd
211 rSerGlyHisPheGlnGlnLeuLysArgLysAspHisGlnLeuTyrPargProGlnArgG 231
1437 TTTGCTCGAATATGTTTCGTTCTACGGAATGATGTTTCCAAACTTCTCTTCAACC 1496
Dd
231 yLeuProGluArgSerValSerValLeuArgAlaTrpMetPheGlnAsnPheLeuHisPr 251
1497 TTACCGGAAGATTCGAGAACATCTCTAGCTATACAGAGTGGCTTGACAGAGTCA 1556
Dd
251 cTyrProLysAspAlaGluLysGlnLeuLeuAlaValLysSerGlyLeuThrArgSerGl 271
1557 GGATCAAACTGTTTATAAATCGCGGTTAGGCTATGGAAGCCGATGATAGAAGAGAT 1616
Dd
271 nValSerAsnTrpPheIleAsnAlaArgValArgLeuTyrLysProMetIleGluGluMe 291
1617 GTATCGGAATCAACAGAGGAGCTC 1644
Dd
291 tTyrAlaGluMetAsnArgLysIle 300

RESULT 3

Q7Y0Z7 PRELIMINARY; PRT; 699 AA.
AC Q7Y0Z7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Bell-like homeodomain protein 2.
GN BL2.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.

RA Ron N., Parnis A., Lifschitz E.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF375966; AAP47025.1; -;
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 699 AA; 77662 MW; 535A17C9D5A68362 CRC64;

Alignment Scores:

Pred. No.: 4,42e-34 Length: 699
Score: 487.00 Matches: 152
Percent Similarity: 45.27% Conservative: 87
Best Local Similarity: 28.79% Mismatches: 169
Query Match: 14.83% Indels: 120
Dd: 20
Gaps: 20

US-09-423-575-1 (1-1886) x Q7Y0Z7 (1-699)

QY 289 TTTTCATGAAACCAATGAGCAACAAC-----AACCAACAAC 324
Dd
2 TyrTyrGlnGlyThrSerAspAsnAsnIleGlnAlaAspHisGlnGlnGlnHisAsn 21
QY 325 AACACTTTTGTCTCTGGATAATGTCATGACT-----AACCAAAATCTCTT 372
Dd
22 AsnLeuGlyAsnSerAsnAsnAsnIleGlnThrLeuTyrLeuMetAsnProAsnSerTyr 41
QY 373 CTCATGGATTATTACCTTCAAGAGAAGATTCACTTCATTCCTCAACAATGCTCCATGG 432
Dd
42 MetGlnGlyTyrThrThr----- 48
QY 433 AATACCATCAGATCAGATCTCTACAAATGGTGGCTTTGATATTTTCAATTCATCTG 492
Dd
49 AspThrGlnGlnHisLeuGlnGlnGlnAsnGlnHisGlnLeuLeu----- 64
QY 493 ACTAACAAATACITATCATCTTCTCCAGGCTCTATCGATGTTCAAGATAACCGCAATGTT 552
Dd
65 -----PheLeuAsnSerAlaProAlaGlyAsnAlaLeuSerHisAlaAsnIle 81
QY 553 GAG-----TTCATGGCT 588
Dd
82 GlnHisAlaProLeuGlnGlnGlnHisPheValGlyValProLeuProAlaValSerLeu 101
QY 589 CATCTTGGATCATTTAAGACACTATGATGATCTCTCAAAACAACATGCTGGGTTTGA 648
Dd
102 His-----AspGlnIleAsnHisHis-----GlyLeuLeuGlnArgMetTyrPasnAsnGln 118
QY 649 GCAAATAGTAGTTTCAGGCAATTTTCAGGTGTAGTTGGTCCAAAGTGAACCAATGATGCT 708
Dd
119 AspGlnSerGln-----GlnValIleValProSerSerThrValValSer 133
QY 709 ACATTCGGT-----GAAGAGATTTC-----CCG 732
Dd
134 AlaThrSerCysGlyGlyThrThrAspLeuAlaSerGlnLeuAlaPheGlnArgPro 153
QY 733 TTTCTAATTCG-----AATAAAGAAACAATGAGCTTCA 768
Dd
154 IleValValSerProThrProGlnHisArgGlnGlnGlnGlnGlnGlyGlyLeuSer 173
QY 769 TTGAGCTTTCGATCAGATGTTTCTGATGAATGCTCG-----GAGATAAGTCTTTGT 819
Dd
174 LeuSerLeuSerProGlnGlnGlnGlnIleSerPheAsnAsnAsnIleSer----- 191
QY 820 GCACCTACAAGATTAGCCTCAGAGCAAGCTTTTCGAGCAGCAAGACATTTCTAATAAC 879
Dd
192 -----SerSerSerProArgThrAsnAsn 199
QY 880 GTTGTACTCAAGGT-----TTCTCTCACTTATATTGGCTCAAAATAC 924
Dd
200 ValThrIleArgGlyThrMetAspGlyCysSerSerAsnMetIleLeuGlySerLysTyr 219
QY 925 CTTCACTCTGTTCAGAAATACATCTCATTTCCGCCCACTACGCTCGATTTATCTATCT 984
Dd
220 LeuLysAlaAlaGlnGluLeuLeuAspGlu---ValValAsnIleValGlyLysSerAsn 238
QY 985 CGAGGAACCGAGTCAGGAGCTGCTAGTTTCAGCCTTACTTTCACGCTTTTGAGATATAACT 1044

```

Db 239 LysGlyAspAspGlnLysLysAspAsnSer-----MetAsnLysGluLeuLeuPro 255
QY 1045 GAGTTCTTCGATGGTGATCTTAATCACTCGGAGCGGGTTCCGA----- 1089
Db 256 LeuValSerAspValAsnThrAsnSerGlyGlyGlyGluSerSerArg 275
QY 1090 -----TCTCATTTCAAAGGAGAGCATTAGAACCAAG 1122
Db 276 GlnLysAsnGluValAlaIleGluLeuThrAlaGlnArgGlnGluMetLys 295
QY 1123 AAAACCATCTCTTGATCTTCTCAATGCGGATGATCATAGTCATTCGCTAGAT 1182
Db 296 LysAlaLysLeuLeuAlaMetLeuGluValGluGlnArgGlyArgGlnTyrHis 315
QY 1183 GAGATTTCATACGTTATATACCGTTTCCATGCTGCAACCGAGTTA---GATCCACAGTTA 1239
Db 316 GlnMetGlnIleValSerSerPheGluGlnValAlaGlyValGlySerAlaLysSer 335
QY 1240 CACACCGGTTTCCCTCCAAACCGTTCTCTTATACAAAGACCTGAGAGAGAAATC 1299
Db 336 TyrThrGlnLeuAlaLeuHisAlaIleSerLysGlnPheArgCysLeuLysAspAla 355
QY 1300 TCACAGAACATATCTCT-----ATGGGATCTGTA 1329
Db 356 SerGluGlnValLysAlaThrSerLysSerLeuGlyGluAspGluGlyLeuGlyLys 375
QY 1330 TTGGAGAGAGGCAAGACAAAGACTCAAGAAACCTCTATGTTCCACGACATTGCTTCTT 1389
Db 376 IleGluGlySerArgLeuLysPheValAspHisLeuArgGlnGlnArgAla---Leu 394
QY 1390 CAGCAGCTAAAGAAAGAACCATCAGATTGGAGACCTCAACAGAGTTTCCTGAGAA 1449
Db 395 GlnGlnLeuGlyMetMetGlnProAsnAlaTrpArgProGlnArgGlyLeuProGlu 414
QY 1450 TCTGTTTCGGTCTACGGAATGGATGTTCCAAACCTCTTCCACCTTACCCGAAAGAT 1509
Db 415 AlaValSerValLeuArgAlaTrpLeuPheGluHisPheLeuHisProTyrProLys 434
QY 1510 TGGAGAACATCTCTTACGATACGAGTGGCTGACAGAGAGTCAAGTATCAACTGG 1569
Db 435 SerAspLysIleMetLeuAlaLysGlnThrGlyLeuThrArgSerGlnValSerAsn 454
QY 1570 TTTATAAATGCGGGTTAGGCTATGGAAGCGATGATAGAGAGATGATGCGGAAATG 1629
Db 455 PheIleAsnAlaArgValArgLeuTrpLysProMetValGluGluMetTyrLeuGlu 474
QY 1630 AACAAGAGAGAGCTCAATAACAGT 1653
Db 475 ValLysAsnGlnGluGlnAsnSer 482

RESULT 4
Q8LLE4 PRELIMINARY; PRT; 688 AA.
AC Q8LLE4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BEL1-related homeotic protein 5.
GN BEL5.
OS Solanum tuberosum (Potatoe).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]_
RP SEQUENCE FROM N.A.
RA CHEN H., ROSIN F.M., HANNAPEL D.J.;
RT "A KNOX protein of potato interacts with several members of the TALE
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AF406697; AAN03621.1;

```

```

DR GO:0005634; C:nucleus; IEA.
DR GO:0003700; F:transcription factor activity; IEA.
DR GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR006583; POX.
DR Pfam: PF00046; homeobox; 1.
DR ProDom: PD00010; Homeobox; 1.
DR SMART: SM00389; HOX; 1.
DR SMART: SM00574; POX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
SQ SEQUENCE 688 AA; 76459 MW; 4451807DDF525781 CRC64;

Alignment Scores:
Pred. No.: 3,37e-33 Length: 688
Score: 477.00 Matches: 150
Percent Similarity: 45.61% Conservative: 89
Best Local Similarity: 28.63% Mismatches: 170
Query Match: 14.52% Indels: 115
DB: 10 Gaps: 18

US-09-423-575-1 (1-1886) x Q8LLE4 (1-688)
QY 264 TTCTATCAGAAGAAGAGGATTAAGTTTCATAGAAACCAATGAGCAACAACAACAACA 323
Db 2 TyrTyGlnGlyThrSerAspAsnThrAsnIleGlnAlaAspHisGlnArgHis 21
QY 324 CAACACATTTAGTCTCTCGATATCATGACT-----AACCAAAATCTCT 371
Db 21 nHisGlyAsnSerAsnAsnAsnAsnIleGlnThrLeuTyrLeuMetAsnProAsnAsn 41
QY 372 TCTCATGGATTTATACCTTCAGAGAGATTCATCTTCAACATGCTCCACG 431
Db 41 rMetGlnGlyTyrThrThrSer-----AspThrGlnGlnGlnGlnLeuLeu 57
QY 432 GAATACCATCAGATCAGATCCTCTACAAATGGGTGGCTTTGATATTTCAATTCTATGCT 491
Db 57 ----- 57
QY 492 GACTAACAAATACTTATCATCTCTCCACGGTCTCATGATGTTCAAGATAACGCAATCT 551
Db 58 -----PheLeuAsnSerSerProAlaAlaSerAsnAlaLeuCysHisAlaAsn 74
QY 552 TGAG-----TTCATGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 587
Db 74 eGlnHisAlaProLeuGlnGlnGlnHisPheValGlyValProLeuProAlaValSer 94
QY 588 TCATCTCTTGGATCATTTAAGACACTATGATGATCTCTCAACAACATGCTGGGTTTGA 647
Db 94 uHis-----AspGlnIleAsnHisHis-----GlyLeuLeuGlnArgMetTrpAsnAsn 111
QY 648 AGCAATATGATGAGTTCAGGCATTTTCAGGTGTAGTGTGTCAGAGTGAACCAATGATGC 707
Db 111 nAspGlnSerGln-----GlnValIleValProSerSerThrGlyValSer 126
QY 708 TACATTCGCT-----GAAGAAGATTTCCCGTTTCTTAATTCGAATAAAGA-- 753
Db 126 rAlaThrSerCysGlyGlyIleThrThrAspLeuAlaSerGlnLeuAlaPheGlnArgPr 146
QY 754 -----AACATGAGCTTTTCATTCAGTCT 776
Db 146 oIleProThrProGlnHisArgGlnGlnGlnGlnGlnGlnGlyLeuSerLeuSerLeu 166
QY 777 TGCATCAGATGTTTCTGATGAATGCTCG-----GAGATAAGTCTTTTGTGCAGCTAC 827
Db 166 uSerProGlnLeuGlnGlnGlnIleSerPheAsnAsnAsnIleSer----- 181
QY 828 AAGATTAGCTCAGACAGCTCTTCGACGACGACAAAGACATTTCTATAAGCTGTTTAC 887
Db 182 -----SerSerSerProArgThrAsnAsnValThr 192
QY 888 TCAAGGT-----TTCTCTCAACTTATATTTGGCTCAAAATACCTTCACCTC 932
Db ----- 932

```

```

Db 192 eArgGlyThrLeuAspGlySerSerSerAsnMetValLeuGlySerLysTyrLeuLysAl 212
QY 933 TGTTCAGAAATACATCTATCTCATTTCCGCCGATCTCGCTCGATTATTCATCTCGAGAAC 992
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 212 aAlaGlnGluLeuLeuAspGlu---ValValaSnIleValGlyLysSerIleLysGlyAs 231
QY 993 CGAGTCAGGAGCTGCTAGTCTACGCTTTACTTTCACGTTTGTGAGATATAAAGTGAATTTCT 1052
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 231 pAspGlnLysLysAspAsnSer-----MetAsnLysGluSerMetProLeuAlaSe 248
QY 1053 TGATGTGTATTCTAATACTCGAGCGGGTTCGGATCT----- 1092
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 248 aAspValaSnThrAsnSerSerGlyGlyGluSerSerAspGlnLysAsnGluVa 268
QY 1093 -----ACATTTCAAGAGAGACATTAGAGCAAGAAAGAAACCATCTCTT 1136
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 268 lAlaValGluLeuThrThrAlaGlnArgGlnGluLeuGlnMetLysLysAlaLysLeuLe 288
QY 1137 GGATCTCTTCAAAATGGTGGATCATCATATAGTCATTGCGTAGATGATCATACCGT 1196
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 288 uAlaMetLeuGluGluValGluGlnArgTyrArgGlnTyrHisHisGlnMetGlnIleIl 308
QY 1197 TATATCAGCGTTCATCTGTCGACCGAGTTA---GATCCACAGTTACACACCGGTTTGC 1253
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 308 eValLeuSerPheGluGlnValAlaGlyIleGlySerAlaLysSerTyrThrGlnLeuAl 328
QY 1254 CTTCCAAACCGTTTCTCTTATACAAAGACCTGAGAGAGAACTGCAAGARGATAAT 1313
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 328 aLeuHisAlaIleSerLysGlnPheArgCysLeuLysAspAlaIleAlaGluGlnVally 348
QY 1314 CTCT-----ATGGGATCTGTATTGGAGAGAGGCAA 1343
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 348 sAlaThrSerLysSerLeuGlyGluGluGlyLeuGlyLysIleGluGlySerAr 368
QY 1344 AGACAGAGCTCAAGAACTCTATGTTCCACGAGCATTCCTTCTTCAGCAGCTGAACG 1403
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 368 gLeuLysPheValAspHisHisLeuArgGlnGlnArgAla---LeuGlnGlnIleGlyMe 387
QY 1404 AAGAACCACATCAGATTGGAGACCTCAACAGGTTTGCCTGAGAAATCTGTTTCGGTCT 1463
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 387 tMetGlnProAsnAlaTyrArgProGlnArgGlyLeuProGluArgAlaValSerValle 407
QY 1464 AGGAAATGGATGTTCCAAACTCTTCCACCTTACCCGAAAGATTTCGGAGAAACATCT 1523
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 407 uArgAlaTyrLeuPheGluHisPheLeuHisProTyrProLysAspSerAspLysIleMe 427
QY 1524 TCTAGCTATACGAAGTGGCTTCACAAAGATCAGGTATCAAACTGGTTTATAAATGCGCG 1583
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 427 tLeuAlaLysGlnThrGlyLeuThrArgSerGlnValSerAsnTyrPheIleAsnAlaAr 447
QY 1584 GGTTAGCTATGGAAGCCGATGATAGAGAGATGTATCGGAAATGAACAAGAGGAAGCT 1643
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 447 gValArgLeuTyrLysProMetValGluGluMetTyrLeuGluGluValLysAsnGlnGl 467
QY 1644 CAATAACACT 1653
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 467 uGlnAsnSer 470

```

RESULT 5

```

Q94KL4
ID Q94KL4 PRELIMINARY; PRT; 680 AA.
AC Q94KL4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BEL1-like homeodomain 1.
GN BLH1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; rosids;
OC eurooids 1; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID:3702;
[1]

```

```

RP SEQUENCE FROM N.A.
RA Pickowich M.S., Samach A., Modrusan Z., Haughn G.W.;
RT "A family of BEL1-like homeodomain (BLH) proteins in Arabidopsis
thaliana."
RL Submitted (PEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF353034; AAK43836.1; -
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003700; F:transcription factor activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR006563; POX.
DR Pfam: PF00046; homeobox; 1.
DR ProDom: PD00010; Homeobox; 1.
DR SMART: SM00359; HOX; 1.
DR SMART: SM00574; POX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 680 AA; 74457 MW; AA9E845082F5799C CRC64;

```

Alignment Scores:

```

Pred. No.: 9.28e-33 Length: 680
Score: 472.00 Matches: 152
Percent Similarity: 42.01% Conservative: 74
Best Local Similarity: 28.25% Mismatches: 193
Query Match: 14.37% Indels: 119
DB: 10 Gaps: 13

```

US-09-423-575-1 (1-1886) x Q94KL4 (1-680)

```

QY 283 ATAAAGTTTCATAGAAACCCCAATGGACACACACACACACACACTTTTGTCTCTG 342
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 32 ValGlnTyrThrGlnGlnAspAspSerAsnAsnAsnAsnAsnSerAsnSerAsn 51
QY 343 GAT-----AATGTCATGACTTAACCAAAATCTCTTCATGGATTTTATACCTTCA 393
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 52 AsnAsnAsnThrAsnThrAsnThrAsnAsnAsnSerSerPheValPheLeuAspSer 71
QY 394 AGAGAAGATTCAACTTCTCAACAATGCTTCATGGAATACCATCATCATCATCATCT 453
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 72 HisAlaPro-GlnProAsnAlaSerGlnGlnPheValGlyLeuProLeu----- 87
QY 454 CTACAAATGGGTGGCTTTGATATATTTCAATCTATGCTGACTTAACAAATACTTATCT 513
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 88 -----SerGlyHis-- 90
QY 514 TCTCCCGCTTCATCGATGTTCAAGATAACCGCAATGTTGAGTTCATGGCTCTCTCTCT 573
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 91 GluAlaAlaSerIleThrAlaAlaAspAsnIleSerValLeuHisGlyTyrProProArg 110
QY 574 CATCTCTCCATCTCATCTCTTGGATCATTTAAGACACTATGATGATCTCTCAACAAC 633
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 111 ValGlnTyrSerLeuTyrGly-----SerHisGlnValaAspProThrHisGln 126
QY 534 ATGTGGGGTTTGAAGCAAAATAGTAGTTTCAGGCATTTTCAGGTAGTTGTCCTCAAGT 693
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 127 GlnAlaAlaCysGluThrProArgAlaGlnGlnGlyLeuSerLeuThrLeuSerSerGln 146
QY 694 GAACCAATGATGTCTACATTCGGTGAAGAAGATTTCCTGTTTCTTAATTTTCAATAAAGA 753
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 147 GlnGlnGlnGlnGlnGlnHisHisGlnGlnHisGlnProIleHisValGlyPheGlySer 166
QY 754 AACATGAGCTTTTCATGAGTCTTGATTCAGATTCAGATGTTTCTGATGAATGCTCGGAGATAAGT 813
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 167 GlyHisGlyGluAspIleArgValGlySer----- 176
QY 814 CTTTGTGCAGCTACAAGATTAGCTCTCAGAGCAAGCTTCTTTCAGCAGCAAGACATTCT 873
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 177 -----GlySerThr 179
QY 874 AATAACCTTTGTTACTCAAGTTTCTCTCAACTTATATTTGGCTCAAAATACCTTCACTCT 933
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
Db 180 GlySerGlyValThrAsnGlyIleAlaAsnLeuVal---SerSerLysTyrLeuLysAla 198

```

QY 934 GTTCAGAACTACTATCTCATTTCCCGCATACTCGCTCGATTATTCATCTCGAGAAC 993
 DB 199 AlaGlnGluLeuLeuAspGluValValAsnAlaAspSerAsp 212
 QY 994 GAGTCAGGAGCTGAGTTCACCTTTTACCTTTCACCTTTTTCAGAAATATA---ACTGAGTTT 1050
 DB 213 AspMetAsnAlaSerGlnLeuPheSerSerLysLysGlySerCysGlyAsnAspLys 232
 QY 1051 CTTGATGGTATCTTAATACTCGGAGCGGGTTTCGGATCT 1092
 DB 233 ProValGlyGluSerSerAlaGlyAlaGlyGlyGluGlySerGlyGlyValAlaGluAla 252
 QY 1093 -----ACATTCAAGAGAGAGAGATTAAGACCAAGAAA 1125
 DB 253 AlaGlyLysArgProValGluLeuGlyThrAlaGluArgGlnGluLeuMetLysLys 272
 QY 1126 ACCATCTCTTGATCTTCTTCAAGTGGTGAATGATATGATCATTCGCTAGATGAG 1185
 DB 273 AlaLysLeuSerAsnMetLysGluValGluGlnArgTyrArgGlnTyrHisGlnGln 292
 QY 1186 ATTCATACAGTTATATCAGCTTCCATGCTGCAACCGAGTTA---GATCCACAGTTACAC 1242
 DB 293 MetGlnMetValLysSerPheGluGlnAlaAlaGlyGlySerAlaLysSerTyr 312
 QY 1243 ACCCGTTTTCCTCCAAACCGTTTCTTCTTATACAAGAACCTGAGAGAGATTCGC 1302
 DB 313 ThrSerLeuAlaLeuLysThrLysSerArgGlnPheArgCysLeuLysGluAlaAla 332
 QY 1303 AAGAATATCTCTATGGATCTGTATGGAGAGAGCAAGACAACTCAAGAAC 1362
 DB 333 -----GlyGlnLysAlaAlaAsnLysSerLeuGlyGluGluAsp 346
 QY 1363 TCTATG-----TTCCACGACCATTCGCTT 1386
 DB 347 SerValSerGlyValGlyArgPheGluGlySerArgLeuLysPheValAspHisLeu 366
 QY 1387 CTTGACGAG-----CTGAACGAAGAACCATCAGATTGG 1422
 DB 367 ArgGlnGlnArgAlaLeuGlnGlnLeuGlyMetIleGlnHisProSerAsnAlaTyr 386
 QY 1423 AGACCTCAACGAGTTTGCTCAGAAATCTGTTTCGGTCTACGGAATGGATGTCAC 1482
 DB 387 ArgProGlnArgGlyLeuProGlnArgAlaValSerValLeuAlaGlyAlaTyrPheGlu 406
 QY 1483 AACTTCTCTTCACTTACCCGAAAGATTGGGAGAAATCTTCTAGCTATACGAATGGC 1542
 DB 407 HisPheLeuHisProTyrProLysAspSerAspLysHisMetLeuAlaLysGlnThrGly 426
 QY 1543 TTGACAGAGTCAGTATCAACTGTTTATATGCGCGGTAGGCTATGGAAGCG 1602
 DB 427 LeuThrArgSerGlnValSerAsnTyrPheIleAsnAlaArgValArgLeuTyrPhePro 446
 QY 1603 ATGATAGAGAGATGTATCGGAAATGAACAGAGAGAGCAATCAATACAGTCACATTC 1662
 DB 447 MetValGluGluMetTyrMetGluGluMetLysGluGlnAlaLysAsnMetGlySerMet 466
 QY 1663 CCCAAGCGNCA-----ACTCTTCGATG 1686
 DB 467 GluLysThrProLeuAspGlnSerAsnGluAspSerAlaSerLysSerThrSerAsnGln 486
 QY 1687 CCAAAATCTGTTATGATGACCAAGCAATGATCAATAAAGCAACAAT 1734
 DB 487 GluLysSerProMetAlaAspThrAsnTyrHisMetAsnProAsnHis 502
 RESULT 6
 Q9SJ56 PRELIMINARY; PRT; 680 AA.
 ID Q9SJ56
 AC Q9SJ56;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Putative homeodomain transcription factor (At2g35940).

GN AT2G35940.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eumids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., Vanaken S.E., Unayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana";
 RL Nature 402:761-768 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shim P., Southwick A., Shinozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene At2g35940 (GI:15227535).";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shim P., Southwick A., Shinozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones.";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shim P., Southwick A., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
 RA Dale J.M., Hayashizaki Y., Hsuan V.W., Ishida J., Jones T., Kamiya A.,
 RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
 RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Arabidopsis ORF clones.";
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC007017; AAD21463.1; -
 DR EMBL; AY072175; AAL5997.1; -
 DR EMBL; AY096548; AAM20198.1; -
 DR EMBL; AY098554; AAM20705.1; -
 DR EMBL; BT008422; AAP37781.1; -

PIR: H84774; H84774.
HSSP; P01366; IAXH.
GO; GO:000634; C:nucleus; IEA.
GO; GO:0003700; P:transcription factor activity; IEA.
GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
InterPro: IPR001356; Homeobox.
Pfam: PF00046; Homeobox; 1.
ProDom: PD00010; Homeobox; 1.
PROSITE; PS00027; HOMEBOX_1; 1.
PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 680 AA; 74463 MW; 45C0E77956AB8737 CRC64;

Alignment Scores:
Pred. No.: 9.28e-33 Length: 680
Score: 472.00 Matches: 152
Percent Similarity: 42.01% Conservative: 74
Best Local Similarity: 28.25% Mismatches: 193
Query Match: 14.37% Indels: 119
DB: 1.0 Gaps: 13

US-09-423-575-1 (1-1886) x Q9SU56 (1-680)

QY	283	ATAAGTTTCATAGAAACCCCAATGGACAACAACAACAACAACACTTTTAGTTCTCTG	342
Db	32	ValGlnTyrThrGlnGlnAspSerAsnAsnAsnAsnSerAsnAsnSerAsn	51
QY	343	GAT-----AATGTCATCACTAACCAAAATCCTCTTCATGGATTTATACCTTCA	393
Db	52	AsnAsnAsnThrAsnThrAsnThrAsnAsnAsnAsnSerSerPheValPheLeuAspSer	71
QY	394	AGAGAAATTCACCTTCATTCACCAATGCTTCATGGAAATACCATTCAGATCAGATCCT	453
Db	72	HisAlaPro-GlnProAsnAlaSerGlnGlnPheValGlyLeuProLeu	87
QY	454	CTACAATGGTGGCTTGATATTTTCAATCTATGCTAGCTACAAATACTTATCATCT	513
Db	88	-----SerGlyHis--	90
QY	514	TCGCCAGGCTATCGATGTTCAAGATAACCGCAATGTGAGTTGATCGCTCCTCCTCT	573
Db	91	GlualaAlaSerIleThrAlaAlaAspAsnIleSerValLeuHisGlyTyrProProArg	110
QY	574	CATCCTCCTCCATTCATCCCTTTGGATCATTTAAGACACTATGATGTTCTCTCAACAAC	633
Db	111	ValGlnTyrSerLeuTyrGly-----SerHisGlnValAspProThrHisGln	126
QY	634	ATGTGGGGTTTGAAGCAAAATAGTAGTTCAGGCATTTTCAGGTAGTGTGGTCCAAGT	693
Db	127	GlnAlaAlaCysGluThrProArgAlaGlnGlnGlyLeuSerLeuThrLeuSerSerGln	146
QY	694	GAACCAATATGCTACATTCGGTGAAGAAGATTTCCTCGTTTCAATTTCGAATAAAGA	753
Db	147	GlnGlnGlnGlnGlnHisHisGlnGlnHisGlnProIleHisValGlyPheGlySer	166
QY	754	AACAATCAGCTTTCATTGAGTCITGCAATCAGATGTTTCTGATGAATCCTCGAGATAAGT	813
Db	167	GlyHisGlyGluAspIleArgValGlySer-----	176
QY	814	CTTTGTGCAGCTACAAGATTAGCCTCAGAGCAAGCTTCTTGACAGCAAGACATTCT	873
Db	177	-----GlySerThr	179
QY	874	AATAAGTTTACTCAAGTTTCTCTCAACTTATATTTGGCTCAAAATACCTTCACTCT	933
Db	180	GlySerGlyValThrAsnGlyIleAlaAsnLeuVal---SerSerLysTyrLeuLysAla	198
QY	934	GTTCAAGAAATACCTATCTCATTTTCGGCGCATCTCGCTCATTTATTCATCTCGAGGAACC	993
Db	199	AlaGlnGluLeuLeuAspGluValValAsnAlaAspSerAsp	212
QY	994	GAGTCAGGAGCTGCTAGTTCAGCGCTTTCATTCAGCTTTTGGAGATATA---ACTGAGTTT	1050

QY 1405 AAGAACCATCAGATTGGAGACTCAACGAGGTTGCCCTGAGAAATCTTTCGGTCTCA 1486

Db 41 GlnGlnGlnGlnSerSerValProLeuSerSerLysLysTyrMetLysAlaAla 60
QY 937 CAAGAAATATCTATCTATTCGCGCATCTCGCTCGATTAATTCATCTCGAGAACCGAG 996
Db 61 GlnGlnLeuLeuAspGlu-----ValValAsnValGlyLysSerMetLysSerThrAsn 78
QY 997 TCA---GGAGCTGTAGTTCAGCTTTTACCTTCACGTTTGAGATATACTAGTTT---1050
Db 79 SerThrAspValValAsnAsnAspValLysSerLysAsnMetGlyAspMetAsp 98
QY 1051 -----CTTGATGGT-----GATTCCTAACTCGGAGCGGGTTTCGGATCTACA 1095
Db 99 GlyGlnLeuAspGlyValGlyAlaAspLysAspGlyAlaProThrThrGluLeuSerThr 118
QY 1096 TTTCAGAGGAGACATTAGAGCAAGCAAAACCCATCTCTTGATCTTCTTCAATGCGTG 1155
Db 119 GlyGluArgGlnGlnMetLysLysAlaLysLeuValAsnMetLeuAspGluVal 138
QY 1156 GATGATCGATATAGTTCATTCGCTAGATGAGATTCATACGGTTTATATCAGGGTTCCATGCT 1215
Db 139 GluGlnArgTyrArgHisTyrHisHisGlnMetGlnSerValHisHisTrpLeuGluGln 158
QY 1216 GCACACGAGTAA---GATCCACAGTTACACCCCGTTTGCCTCCAAACCGTTTCTTC 1272
Db 159 AlaAlaGlyIleGlySerAlaLysThrTyrThrAlaLeuAlaLeuGlnThrIleSerLys 178
QY 1273 TTATACAGAACCTGAGAGAGAAATCTGCAAGAGATA-----1311
Db 179 GlnPheArgCysLeuLysAspAlaIleIleGlyGlnIleArgSerAlaSerGlnThrLeu 198
QY 1312 -----ATCTCTATGGATCTGTATTGGAGAGAGCAAGAGCAAGACTCAAGAAACC 1362
Db 199 GlyGluGluAspSerLeuGlyGlyLysIleGlyLysArgLysArgLysPheValAspAsn 218
QY 1363 TCTATGTTCCACAGCATTCCTTCTTCAGCAGCTGAAACGAGAAACCATCAGATTGG 1422
Db 219 GlnLeuArgGlnGlnArgAla---LeuGlnGlnLeuGlyMetIleGlnHisAsnAlaTrp 237
QY 1423 AGACCTCAACGAGTTTGCCTGAGAAATCTGTTTCGTTCTACGGAATGGATGTCCAA 1482
Db 238 ArgProGlnArgGlyLeuProGluArgAlaValSerValLeuArgAlaTrpLeuPheGlu 257
QY 1483 AACTTCCTTCACCTTACCGGAAAGATTGGGAGAACATCTCTAGCTATACAGATGGC 1542
Db 258 HisPheLeuHisProTyrProLysAspSerAspLysMetMetLeuAlaLysGlnThrGly 277
QY 1543 TTGACAGAGAGTCAGGTATCAAACTGTTTATAATGCGCGGTTAGGCTATGGAAGCG 1602
Db 278 LeuThrArgSerGlnValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysPro 297
QY 1603 ATGATAGAGAGATGATCGGAATGAAACAGAGGAGAGCTCAATAAC 1650
Db 298 MetValGluGluMetTyrLeuGluGluIleLysGluHisGluGlnAsn 313

RESULT 9

QY028
ID QY028 PRELIMINARY; PRT; 393 AA.
AC QY028;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Bell-like homeodomain protein 1 (Fragment).
GN BL1
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]_"
RP SEQUENCE FROM N.A.
RA Ron N.; Parnis A.; Lifschitz E.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF375965; AAP47024.1; --
KW DNA-binding; Homeobox; Nuclear protein.
FT NON_TER 1
SQ SEQUENCE 393 AA; 44877 MW; D4A45EADA37F084 CRC64;
Alignment Scores:
Pred. No.: 6,38e-31 Length: 393
Score: 450.50 Matches: 132
Percent Similarity: 45.87% Conservative: 57
Best Local Similarity: 32.04% Mismatches: 152
Query Match: 13.72% Indels: 71
DB: 10 Gaps: 13
US-09-423-575-1 (1-1886) x QY028 (1-393)
QY 469 TTGTATATTTTCAATTTCTATCTGCTACTAACAATACTTATCATCTTCTCCACGGTCTATC 528
Db 3 PheAspMetTyrGlnSerAspThrThrAlaTyrGlnProHisGlyGlyLeuSer--- 21
QY 529 GATGTTCAAGATAACCGCAATTTGAGTTTCATGCTCTCTCTCTCATCTCTCCCTCCACTT 588
Db 22 -----ArgSerIleGluPheValAsnHisProAspPheThrThrAspSer 36
QY 589 CATCTTTGGATCAATTTAAGACACTATGATGATTCCTCAAAACAACATGCGGGTTTGA 648
Db 37 HisAspValAsnHisSerArgHisLeuMetAspLeuLeuGlyAlaSerHisAspAlaAsn 56
QY 649 GCAATAGTAGTTTCAGGCATTTTCAGGTGATGTTGTCAGTGAACCAATGATGCT 708
Db 57 ThrAsnGlnGlnAlaGlnArgLeuSerLeuGly---SerHisSerLeuValSer 75
QY 709 ACATTTCGGTGAAGAAGATTTCCTGTTCTTAATTTTCG-----AATAAAGAAACAAT 759
Db 76 ThrPheThrAsnAsn-----ProSerTyrMetAsnGlnGluIleAspGlnArgAsnAsn 93
QY 760 GAGCTTTCAATGAGTTCATGATCAGATGTTTCTGATGAATCTCGAGATAAGTCTTTGT 819
Db 94 GluPheSerPheSer-----98
QY 820 GCAGCTACAGATTAAGCTCAGACAAAGCTTCTTCAGCAGCAAGACATTTCTAATAAC 879
Db 99 -----AlaAlaAlaMetAsnGlnSerPheSerAsnVal 109
QY 880 GTTGTACTCAAGTTTCTCTCAACTTATATTTGGCTCAAAATACCTCTCTGTTCAA 939
Db 110 CysGlyThrGluSerPheValSerAlaIleGlyAsnSerLysTyrLeuLysProThrGln 129
QY 940 GAAATACTATCTCATTTTCGCGGCATAC-----TCGCTCGATTATTC 981
Db 130 SerLeuLeuGluGluLeuValCysIleGlyLysThrIleAspSerSerAsnGluLys 149
QY 982 -----TCTCGAGGAACCGAGTCTAGGAGCTGCTAGTTTCAGCTTACTTCA 1026
Db 150 PheIleArgArgLeuSerArgAsnSerLysLysGlySerLeuSer-----164
QY 1027 CGTTTTGAGAATATACTAGTTTCTTGATGTTGATTTCTAATACTCGGAGCGGGTTTC 1086
Db 165 -----LeuArgAlaMetLeuLysGlyGluIleProProAsnAsnGluLeuPhe 180
QY 1087 GGATCTCATTTCAAAGGAGAGCATTAGAAGCAAGAAACCCATCTCTTGGATCTTCTT 1146
Db 181 Asn-----GluArgHisGluLeuTyrValLysIleMetLysLeuIleAlaLeuLeu 197
QY 1147 CAATGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1206
Db 198 GluGluValGluArgArgTyrGluGlnTyrTyrGlnHisMetGluGluValThrSerThr 217
QY 1207 TTCCATCTCTCAACCGAGTTAGATCCA---CAGTTACACACCGGTTTGCCTCCCAACC 1263
Db 218 PheGluValIleAlaGlyPheGlyAlaGlyAlaTyrThrAlaLeuAlaLeuGlnAla 237
QY 1264 GTTTCCTTCTTATACAAAGAACCTGAGAGAGAGATCTGCAAGAGAGATAATCTCTATGGGA 1323
Db 1323

Db 238 MetSerArgHisPheCysCysLeuArgAspSerIleLeuSerGlnIleAsnPhelIleArg 257
 QY 1324 TCTGTATTGGAGAGAGGCAAGACAAAG-----ACTCAAGAAACCTCTATGTTTC 1371
 Db 258 GlnLysMetProArgAspValProLysIleSerSerGlyLeuSerHisLeuSerLeuPhe 277
 QY 1372 CACGACGATGCGCTT-----CTTCAGAGCTGAAA-----CGAAGAAC 1410
 Db 278 GlnLysGluThrLeuGlnAsnArgIleSerLeuGlnGlnLeuGlyIleGlnSerAsn 297
 QY 1411 CATCATGTTGGAGACCTCAACGAGGTTGCTGAGAAATCTGTTTCGTTCTACGGAAT 1470
 Db 298 ArgGlnAlaTrpIlePheArgGlyLeuProGluThrSerValAlaPheLeuArgSer 317
 QY 1471 TCGATGTTCCAAACTCTCTCCACCTTACCCGAAAGATTCGGAGAAACATCTCTAGCT 1530
 Db 318 TrpLeuPheGluHisPheLeuHisProTyrProAsnAspSerGluLysLeuMetLeuSer 337
 QY 1531 ATACGAGTGGCTTGACAGAGAGTCAGGTATCAAACTGGTTTATAAATCGCGGGTTAGG 1590
 Db 338 SerGlnThrGlyLeuSerLysAsnGlnValSerAsnTrpPheIleAsnAlaArgValArg 357
 QY 1591 CTATGGAACCGCATAGAGAGATGTATGCGGAA 1626
 Db 358 LeuTrpLysProMetIleGluMetTyrLysGlu 369

RESULT 10

Q8LLE1 PRELIMINARY; PRT; 532 AA.
 AC Q8LLE1;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE BEL1-related homeotic protein 14 (Fragment).
 GN BEL14.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Desiree;
 RA Chen H., Rosin F.M., Hannapel D.J.;
 RT "A KNOX protein of potato interacts with several members of the TALE
 RT family of transcription factors."
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP406700; AN03624.1;
 DR GO: GO:0005634; C:nucleus; IEA.
 DR GO: GO:0003700; F:transcription factor activity; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR006563; POX.
 DR Pfam: PF00046; homeobox. 1.
 DR ProDom: PD000010; Homeobox; 1.
 DR SMART: SM00389; HOX; 1.
 DR SMART: SM00574; POX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00071; HOMEBOX_2; 1.
 FT NON TER 1
 SQ SEQUENCE 532 AA; 59757 MW; 54940BFD29B18474 CRC64;

Alignment Scores:

Pred. No.: 6,89e-31 Length: 532
 Score: 450.50 Matches: 132
 Percent Similarity: 46.51% Conservative: 61
 Best Local Similarity: 31.81% Mismatches: 123
 Query Match: 13,72% Indels: 99
 DB: 10 Gaps: 17

US-09-423-575-1 (1-1886) x Q8LLE1 (1-532)

QY 610 CACTATGATGATTCCTCAAAACAATGTGGGGTTTTTCAAGCAAAATAGTGAGTTTCAGGCA 669

Db 31 HisHisAspAspHisGlnGlySer---TrpHisHisAspAsnAsnArgThrLeuLeuVal 49
 QY 670 TTTTCAGGTTAGTGGTCCAAGTGAACCAATGATGCTCTACATTGGTGAAGAGATTTC 729
 Db 50 -----AspAspProSerMetArgCysVal-----Phe 58
 QY 730 CCGTTTCTAATTTCAATAAAGAAATGAGCTTTTCATTGAGTCTTTCATCAGATGTT 789
 Db 59 ProCysGluGlyAsnGlnArgProSerHisGlyLeuSerLeuSerLeuCys----- 75
 QY 790 TCTGATCAATGCTCGAGATAAGTCTTTGTGAGCTTACAGATTAGCTTCAGAGCAGCT 849
 Db 76 SerSerAsnProSerSerIleGlyLeu-----GlnSerPhe 87
 QY 850 TCTTGACGACGACAAACACATTTCTAATAAGTGTGTTTACTCAAGGTTTC----- 897
 Db 88 GluLeuArgHisGlnAspLeuGlnGlyLeuIleHisAspGlyPheLeuGlyLysSer 107
 QY 898 -----TCTCAACTTATATTTGGCTCAAAATACCTT 927
 Db 108 ThrAsnIleGlnGlyTyrPheHisHisGlnValArgAspSerLysTyrLeu 127
 QY 928 CACTCTGTTCAAGAAATACTATCTCTATTCGCGCATACTCGCTC-----GAT 975
 Db 128 GlyProAlaGlnLeuLeuSerGluPheCysSerLeuGlyIleLysLysAsnAsp 147
 QY 976 TATTCTATCT-----CGAGAACCGAGCTCAGAGCTCTAGTTTCAGCCTTT 1020
 Db 148 HisSerSerLysValLeuLeuLysGlnHisGluSerThrAlaSerThrSerLys--- 166
 QY 1021 ACTTCAGTTTTCAGAAATATACTAGTTCCTTCGATGCTGATCTTAATCACTCGGAGCG 1080
 Db 167 -----LysGlnLeuLeuSerLeuLeuAsp----- 174
 QY 1081 GGTTTCGATCTACATTTCAAAGGAGCATTTAGAAGCAAGAAAGAAACCCATCTCTTGGAT 1140
 Db 175 -----LeuLeuGluLeuGlnLysArgLysThrLysLeuLeuGln 187
 QY 1141 CTCTCTCAATGTTGATGATGATATGATGATGATGATGATGATGATGATGATGATGAT 1200
 Db 188 MetLeuGluValAspArgArgTyrLysHisTyrCysAspGlnMetLysAlaVal 207
 QY 1201 TCAGCGTTTCATGCT---GCAACCGAGTTAGATCCACAGTTACACACCCGGTTGCGCTC 1257
 Db 208 SerSerPheGluAlaValAlaGlyAsnGlyAlaAlaThrValThrSerAlaLeuAlaSer 227
 QY 1258 CAAACCGTTTCTCTTATACAAAGAACCTGAGAGAGAGAAATCTGCAAGAGATA----- 1311
 Db 228 ArgAlaMetSerArgHisPheArgCysLeuArgAspGlyLeuValAlaGlnIleLysAla 247
 QY 1312 -----ATCTCTATGGATCTGTTATGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1356
 Db 248 ThrLysMetAlaMetGlyGlu-----LysAspSerThrSerThrLeuLeu 262
 QY 1357 -----GAAACG-----TCTATGTTTCCACAG 1377
 Db 263 ProGlySerThrArgGlyGluThrProArgLeuArgLeuLeuAspGlnThrLeuArgGln 282
 QY 1378 CATTGCGCTTCTTCAGCAGCTGAAACGAAACCATCAGATTGGAGACCTCAACGAGGT 1437
 Db 283 GlnLysAlaPheGlnGlnMetAsnMetMetGluThrHisProTyrArgProGlnArgGly 302
 QY 1438 TTGCTCGAGAAATCTGTTTCGGTTCTACGGAATGATGTTTCCAAACTTCCTTCACCTC 1497
 Db 303 LeuProGluArgSerValSerValLeuArgAlaTrpLeuPheGluHisPheLeuHisPro 322
 QY 1498 TACCCGAAAGATTCGGAGAAACATCTCTTACGCTATACGAAAGTGGCTGACAAAGCTCAG 1557
 Db 323 TyrProSerAspValAspLysHisIleLeuAlaArgGlnThrGlyLeuSerArgSerGln 342
 QY 1558 GTATCAACATGGTTTATAAATGCGCGGTTAGGCTATGGAAGCGGATGATAGAGAGATG 1617

Db 343 ValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetValGluGluMet 362
QY 1618 TATCGGGAATGAACAGAGAGAGCTCAATAACAGTACACATTCACCAACGAGCAACT 1677
Db 363 TyrLeuGluGluThrLysGluGluAsnValGly-----SerProAspGly----- 378
QY 1678 CTTCCGAATGCCAAATCTGTATGATGAGCCCAAGCAATGATGATAA 1722
Db 379 -----SerLysAlaLeuIleAspMetThrIleHisGln 390
RESULT 11
Q9FWS9 PRELIMINARY; PRT; 524 AA.
AC Q9FWS9; 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE F1616.6 protein (BELL1-like homeodomain 3) (Homeodomain protein BELL1, putative).
DE F1616.6 OR BLH3
GN Arabidopsis thaliana (Mouse-ear cross).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altari H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Buehler E., Chao Q., Chin C., Chio J., Choi E., Gonzalez A.,
RA Howing B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharbek N., Pham P., Sakano H., Shinn P.,
RA Toriumi M., Vaynsberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Pickovich M.S., Samach A., Modrusan Z., Haughn G.W.;
RA "A family of BELL1-like homeodomain (BLH) proteins in Arabidopsis thaliana";
RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RA "Full-length messenger RNA sequences greatly improve genome annotation";
RL Genome Biol. 0:0-0(2002).
RN [4]
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; ACQ23754; AAG13065.1; -;
DR EMBL; AF353093; AAK43835.1; -;
DR EMBL; AY085278; AAW62510.1; -;
DR F1R; F96784; F96784.
DR HSP; P41778; IUD6.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 524 AA; 59679 MW; 942461E20FB3F39 CRC64;

Alignment Scores:
Pred. No.: 2,58e-30 Length: 524
Score: 444.00 Matches: 163
Percent Similarity: 42.55% Conservative: 74

Best Local Similarity: 29.26% Mismatches: 170
Query Match: 13.52% Indels: 150
DB: 10 Gaps: 21
US-09-423-575-1 (1-1886) x Q9FWS9 (1-524)
QY 303 AATGGACACAAACAAACAAACACCTTTAGTTCTCTGGAT----- 345
Db 22 AsnGluGlnGlnGlnGlnGlnGln-AlaSerSerSerSerAlaAlaSerPheSerG1 41
QY 346 -----AATGTCATGACTAACAACAAATCCTCTCTCATGTGATTTATACCTTC 392
Db 41 uileValSerGlyAspValA:GAsnAsnGlu-----MetValPheIleProPr 57
QY 393 AAGAGAAGATTCAACTTCTCAATCTCAACAAATGCTTCCATGGAATACCATCAGATCAGATCC 452
Db 57 OThrSerAspValAlaValAsnGlyAsnVal-----ThrValSerSerAsnAs 73
QY 453 TCTACAAATG-----GGTGGCTTTGATATTTTCAATTCTATGCTGACTAACAAATACTT 506
Db 73 pLeuSerPheHisGlyGlyGlyLeuSerLeuSerLeuGlyAsnGlnIleGlnSerAlaVa 93
QY 507 ATCATCTTCTCCACGGTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTCATGGCTCC 566
Db 93 lserValSerProPheGlnTyr----- 100
QY 567 TCTCTCTCATCTCTCCACTTCTCATCTTGGATCATTTAAGACACTATGATTCCTC 626
Db 101 -----HisTyrGlnAsnLeuSe 106
QY 627 AAACACATGTTGGGTTTGAAGCAATAGTAGTTCAGGCATTTTCAGGTAGTTGG 686
Db 106 rAsnGlnLeuSerTyrAsnAsnLeuAsnPro----- 116
QY 687 TCCAAGTGACCAATGATCTACATTCGGTGAAGAAGATTTCCTCGTTCTTAATTTCGAA 746
Db 117 -----SerThrMetSerAspGlu----- 122
QY 747 TAAAGAAACAAATGAGCTTTCATGCTTTCATGATGATTTCTGATGAATCTCGGA 806
Db 123 -----AsnGlyLysSerLeuSerValHisGlnHisSerAspGlnIleLeuPr 139
QY 807 GATAAGTCTTTGTGCAGCTACAGATTAGCTCAGACCAAGCTTCTTGACACCAAGA 866
Db 139 oSerSerVal-----TyrAsnAs 145
QY 867 CATTTCTAATAACGTTGTT-----ACTCAAGGTTTCTC 899
Db 145 nAsnGlyAsnAsnGlyValGlyPheTyrAsnAsnTyrArgTyrGluThrSerGlyPheVa 165
QY 900 TCACTTATATTTGGCTCAAAATACCTTCACTCTGTTCAAGAATACTATCTCATTTCCG 959
Db 165 lSerSerValLeuArgSerArgTyrLeuLysProThrGlnGlnLeuLeuAspGluValVa 185
QY 960 CGCATACTCGCTCGATTAT-----TCATCTCGAGGAACCGA 995
Db 185 lSerValArgLysAspLeuLysLeuGlyAsnLysLysMetLysAsnAspLysGlyGlnAs 205
QY 996 GTCAGGAGCTGCTAGTTCAGCCTTTTACTTCAGGTTTGGAGATATAACTAGTTCTTGA 1055
Db 205 pPheHisAsnGlySerSer-----AspAsnIleThrGlu-----As 217
QY 1056 TGGTGATTCTAATAACGCGAGCGGGTTCGGATCTACATTTCAAGAGGAGCATTTAGA 1115
Db 217 pAspLysSerGlnSerGlnGlnLeu-----SerProSerGluArgGlnGlnLeuG1 234
QY 1116 AGCAAGAAACCCATCTCTTGGATCTTCTTCAAAATGGTGGATGATGATATATGATTCATTG 1175
Db 234 nSerLysLysSerLysLeuLeuThrMetValAspGluValAspLysArgTyrAsnGlnTy 254
QY 1176 CGTAGATGATTCATACGGTTTATATCAGCGTTTCCATGCTGTCACCGAGTTAGAT---CC 1232
Db 254 rHisHisGlnMetGluAlaLeuAlaSerSerPheGluMetValThrGlyLeuGlyAlaAl 274


```

QY 1357 -----GAAACC-----TCATGTTCCACGACGATTCCTTCTT 1389
Db 245 ArgGlyGluThrProArgLeuArgLeuLeuAspGlnThrLeuArgGlnGlnLysAlaPhe 264
QY 1390 CAGACGCTGAACGAAAGAACCATCAGATTGAGACCTCAACGAGGTTTGCTCGAGAAA 1449
Db 265 GlnGlnMetAsnMetMetGluThrHisProTyrArgProGlnArgGlyLeuProGluArg 284
QY 1450 TCGTTTTCGTTCTACGGAATTCGATGTTCCAAACTTCTTCAACCTTACCGGAAGAT 1509
Db 285 SerValSerValLeuArgAlaThrPheGluHisPheLeuHisProTyrProSerAsp 304
QY 1510 TCGAGAAACATCTTCTAGCTATACCAAGTGGCTTGACAAGAGTCAAGTATCAAACTGG 1569
Db 305 ValAspLysHisIleLeuAlaArgGlnThrGlyLeuSerArgSerGlnValSerAsnTrp 324
QY 1570 TTTATAAATCGCGGGTATGAGCTATGGAAGCCGATGATAGAGAGATGATCGGAATG 1629
Db 325 PheIleAsnAlaArgValArgLeuTrpLysProMetValGluGluMetTyrLeuGluGlu 344
QY 1630 AACAGAGGAAG----- 1641
Db 345 ThrLysGluGluGluGluGluAsnValGlySerGlnAspGlySerLysAlaLeuLeuAsp 364
QY 1642 -----CTCAATAC-----AGTCACATTCAACCAACGACCAACTCTTCA 1683
Db 365 GluMetThrIleAsnAsnHisGlnSerSerHisIleValGlnLysProAsnLeuValArg 384
QY 1684 ATCCCAAAATCTGTTATGATGACCAAGCAATGCAT-----AAATAAGACAAC 1731
Db 385 IleGluSerGluCysIleSerSerIleIleAsnHisProHisAspLysAsnAspGln 404
QY 1732 AAT 1734
Db 405 Asn 405

```

RESULT 15

```

Q94K35 PRELIMINARY; PRT; 759 AA.
ID Q94K35;
DC Q94K35;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Homeodomain protein JUBEL1.
GN JUBEL1.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]_TaxID=4513;
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Bonus;
RA Mueller J., Wang Y., Franzen R., Santi L., Salamini F., Rohde W.;
RT "Interactions between barley TALE homeodomain proteins suggest a role
RT for protein-protein associations in the regulation of Knox gene
RT function."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF334758; AK38645.1; -.
DR GO; GO:0003634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR006563; POX.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00574; POX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
SQ SEQUENCE 759 AA; 80092 MW; 982CE4AB3F4EA2A0 CRC64;

```

Alignment Scores:

```

Pred. No.: 8,11e-29 Length: 759
Score: 427.50 Matches: 130
Percent Similarity: 44.09% Conservative: 75
Best Local Similarity: 27.96% Mismatches: 179
Query Match: 13.02% Indels: 81
DB: 10 Gaps: 13
US-09-423-575-1 (1-1886) x Q94K35 (1-759)
QY 445 TCAGATCCTCTACAAATGGGTGGCTTTGATATTTTCAATTTCTATGCTGACTAAACAAATAC 504
Db 194 SerSerProLysGlnGlnGlyGlyMet-----ATCGATGTTCAAGATAACCGCAATGTT 552
QY 505 TTATCATCTTCTCCACGCTCT-----ATCGATGTTCAAGATAACCGCAATGTT 552
Db 205 LeuIaThrAspProAlaAlaAlaMetGlnLeuPheLeuMetAsnProGlnGlnGln 224
QY 553 GAGTTATGGCTCTCTCCCTCATCTCTCCCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 612
Db 225 GlnSerArgSerSerProThrSerProProSerAspAlaGlnSerAlaGlnHis 244
QY 613 -----TATGATGATCTCTCAACCAACATGTGGGTTTTGAAGCAAT 654
Db 245 HisGluIaPheGlnAlaTyrGlyAsnAlaAlaSerSerPheGlyGlyGlyAlaGly 264
QY 655 ---AGTCAGTTTTCAGGCATTTTCAGGTGTAGTTGGTCCAAAGT-----GAACCAATGAT 705
Db 265 ValValGluGlyGlnGlyLeuSerLeuSerLeuSerProSerLeuGlnGlnGluMet 284
QY 706 TCTACATCGGTGAAGAAGATTCCCGTTTCTAATTTTCGAATAAAGAAACAATGAGGTT 765
Db 285 AlalysGlnAlaGluGlu-----LeuArgValArgAspGlyValLeu 298
QY 766 TCATTGATCTTGCATCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTTTGTCAGCT 825
Db 299 TyrPheAsnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 318
QY 826 ACAAGATTAGCTCAGACAGAGCTTCTTGACAGCAGCAAGACATTTCTAATAACGTTGTT 885
Db 319 ProMetAlaLeuHisGlyGlnValGlySerMetGlyGlnGlnLeuHisValGlyTyrGly 338
QY 886 ACTCAAGGTTTCTCTCAACTATATTTGGTCCAAATACCTTCTCTCTCTCTCTCTCTCTCTCT 945
Db 339 ProAlaGlyValAlaGlyValLeuArgAsnSerLysTyrThrArgAlaAlaGlnGlnLeu 358
QY 946 CTATCTCATTTCCGCCGATCTCGCTCGATTATTCATCTCGAGGACCGAGTCAGAGCT 1005
Db 359 LeuAspGluPheCysSerValGlyArgGlyGlnThrIleLysGlyGlyArgGlyGly 378
QY 1006 GCTAGTTTCAGCTTACTTCTCAGCTTTGAGAATAATACTAGTTTCTTGATGGTGATCT 1065
Db 379 SerSerSer-----AsnProAsnAlaSerLysGlyGlyProSer 391
QY 1066 AATAACTCGGAGCGGGTTCGGATCTACA----- 1095
Db 392 SerSerGlyAlaAlaGlnSerProSerSerAlaSerLysGluProGlnLeuSerPro 411
QY 1096 TTTCAAAGAGAGCATTAGAAGCAAGAAACCCATCTCTTGGATCTTCTTCAAAAGGTG 1155
Db 412 AlaAspArgPheGluGlnGlnArgLysLysAlaLysLeuIleSerMetLeuAspGluVal 431
QY 1156 GATGATCGATATAGTCATTGGGTAGATGAGATTCATACGCTTATATCA----- 1203
Db 432 AspArgArgTyrAsnHisTyrCysAspGlnMetGlnMetValValAsnPheAspSer 451
QY 1204 -----OCGTTTCCATGCTGCAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCTC 1257
Db 452 ValMetGlyPheGlyAlaAlaThrPro-----TyrThrAlaLeuAlaGln 466
QY 1258 CAACCGGTTTCTCTTATACAGAACCTGAGAGAGAGATCTGC----- 1302
Db 467 LysAlaMetSerArgHisPheArgCysLeuLysAspAlaIleAlaAlaGlnLeuArgHis 486

```

QY 1303 -----AAGAGATATCTCTATGGGATCTGTATTCGAGAGAGGCAAA 1344
Db 487 ThrCysGluLeuGlyGlyLysAspAlaGlyThrSerSerGlyLeuThrLysGlyGlu 506
QY 1345 -----GACAAAGACTCAAGAAACCTCTATGTTCCACGAGCAT 1380
Db 507 ThrProArgLeuArgAlaIleAspGlnSerLeuArgGlnArgAlaPheHisMet 526
QY 1381 TCCCTTCTTACACAGCTGAACAAAGAACCATCATGTTGGAGACCTCAACGAGGTTTG 1440
Db 527 GlyMetMetGluGln-----GluAlaTrpArgProGlnArgGlyLeu 540
QY 1441 CCTGAGAAATCGTTTCGTTCTACGGAATGGATGTTCCAAAAGTTCTCTCACCCCTTAC 1500
Db 541 ProGluArgSerValSerIleLeuArgSerTrpLeuPheGluHisPheLeuHisProTyr 560
QY 1501 CCGAAAGATTCCGAGAAACATCTTCTAGCTATACGAGTGGCTTGACAAAGATCAGSTA 1560
Db 561 ProSerAspAlaAspLysHisLeuLeuAlaArgGlnThrGlyLeuSerArgAsnGlnVal 580
QY 1561 TCAAACTGGTTTATAAATCCGCGGTTAGGCTATCGAAGCCGATCATAGAGAGATGTAT 1620
Db 581 SerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluGluMetTyr 600
QY 1621 GCGGAAATGACAAAGAGGAAGCTCAATAACAGTCACTTCAACCCACGACCAACTCTT 1680
Db 601 GlnGlnGluThrLysGluLeuGluGlySerSerAlaAlaAlaGlyGlyGlyVal 620
QY 1681 CGAATGCCAAATCT 1695
Db 621 GlyGlyProGluSer 625

Search completed: September 2, 2004, 12:11:44
Job time : 143 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2004, 05:19:27 ; Search time 4991 Seconds
(without alignments)
11284.326 Million cell updates/sec

Title: US-09-423-575-1
Perfect score: 1886
Sequence: 1 attaggtatataaatgttgc.....ttaaaaaaaaaaaaaaaaaa 1886

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vri:*

28: gb_gss1:*

29: gb_gss2:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	508	26.9	511	9 AW004482	AW004482 701931784
C 2	411.4	21.8	423	9 AI999808	AI999808 701857547
C 3	406.2	21.5	423	9 AV808362	AV808362 AV808362
C 4	388.4	20.6	390	14 CB261894	CB261894 83-E8862-

5	277.6	14.7	773	14	CA917399	CA917399 EST641546
6	274.8	14.6	535	28	BH476676	BH476676 BQGN7677F
7	262.2	13.9	775	13	BQ990079	BQ990079 QGF15G17.
8	256.4	13.6	719	28	BZ033022	BZ033022 oeh45H05.
C 9	251.6	13.3	270	9	AV800635	AV800635 AV800635
10	243	12.9	676	13	BU011811	BU011811 QGJ17D24.
C 11	232.4	11.8	709	28	BZ031456	BZ031456 oef98B09.
12	216.4	11.5	751	13	BU837920	BU837920 T107B12 P
13	147.4	7.8	672	13	BU024863	BU024863 QHF6H21.Y
14	147.4	7.8	698	13	BU026692	BU026692 QHG17J13.
15	146	7.7	328	29	AL946255	AL946255 Arabidops
16	145.8	7.7	657	13	BU027235	BU027235 QHG3D05.Y
17	139.2	7.4	415	13	BQ848536	BQ848536 QCA7H07.Y
18	139.2	7.4	706	13	BQ859673	BQ859673 QGC13K23.
19	135.2	7.4	740	13	BQ867350	BQ867350 QGD10I16.
20	135.2	7.2	777	14	CB343619	CB343619 CA32EN000
21	133.4	7.1	333	12	BG662137	BG662137 Ljirmpesc
22	133.4	7.1	669	13	BQ148285	BQ148285 NF065F11F
23	133.4	7.1	677	12	BG448606	BG448606 NF045H07N
24	133.2	7.1	360	13	BU765815	BU765815 sas20D01.
25	133	7.1	616	14	CA8319352	CA8319352 MCT026E10
26	132.8	7.0	850	14	CB292855	CB292855 UCRC801.0
27	132.6	7.0	393	9	AU294697	AU294697 AU294697
28	132.4	7.0	706	12	B1272808	B1272808 NF092G09F
29	132	7.0	520	12	BM886468	BM886468 sam17a04.
30	132	7.0	555	12	BM520903	BM520903 eal31h09.
31	130.8	6.9	647	14	CD718786	CD718786 VVB154B04
32	130.8	6.9	658	14	CD717521	CD717521 VVB154B07
33	130.8	6.9	661	14	CD719099	CD719099 VVB157E07
34	130.8	6.9	675	14	CD718777	CD718777 VVB154A07
35	130.8	6.9	783	10	AW688895	AW688895 NF004E06S
36	130.2	6.9	445	10	AW596251	AW596251 s301B04.Y
37	130.2	6.9	479	9	AW102248	AW102248 sd85C07.Y
C 38	130	6.9	138	29	AL756855	AL756855 Arabidops
39	130	6.9	770	14	CA993653	CA993653 PAR09B07
40	129	6.8	570	14	CD056054	CD056054 HO11A03S
41	128.8	6.8	537	12	BM554132	BM554132 sam70e12.
42	128.8	6.8	635	12	BM360047	BM360047 GA_sa002
C 43	128.6	6.8	673	28	BH556826	BH556826 BOHQJ62TF
44	128.6	6.8	480	13	BQ509372	BQ509372 EST616787
C 45	128.6	6.8	513	14	CD662087	CD662087 3529_1_13

ALIGNMENTS

RESULT 1
AW004482/c 511 bp mRNA linear EST 08-SEP-1999
LOCUS AW004482 701931784 A. thaliana, mixed source Arabidopsis thaliana cDNA clone
DEFINITION AW004482 701931784, mRNA sequence.
ACCESSION AW004482
VERSION AW004482.1 GI:5851511
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryote, viridiplantae, Streptophyta, Embryophyta, Tracheophyta, Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, rosids, eurosids II, Brassicales, Brassicaceae, Arabidopsi
1 (bases 1 to 511)
Chen, J., Momiya, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D., Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobrega, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.
Arabidopsis thaliana Gene Expression MicroArray
JOURNAL Unpublished (1999)
CONTACT: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA

Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.

Carpio, T., Policky, J., Suzuki, G., Argentin, C., Shah, S.,
Nobrega, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and
Hanson, D.
Arabidopsis thaliana Gene Expression MicroArray
Unpublished (1999)

TITLE JOURNAL COMMENT

Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.

FEATURES source

1. 423

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/cultivar="Columbia Col-0"

/db_xref="taxon:3702"

/clone="701557547"

/tissue_type="rosette"

/dev_stage="4 - 7 weeks"

/clone_lib="A. thaliana, Columbia Col-0, rosette-3"

/note="Vector: pSPORT; Site 1: NotI; Site 2: SalI; cDNA library was derived from untreated rosette tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was blunt-ended, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."

ORIGIN

Query Match 21.8%; Score 411.4; DB 9; Length 423;
Best Local Similarity 99.5%; Pred. No. 6,1e-68;
Matches 412; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1400 AACGAAAGAACCATCAGATTGGAGACCTCAACGAGGTTGGCTCGAGAAATCTGTTTCGG 1459

DB 414 ACGAAAGAACCATCAGATTGGAGACCTCAACGAGGTTGGCTCGAGAAATCTGTTTCGG 355

QY 1460 TTCTAGGAAATGGATGTTCCAAACTCTTCCACCTTACCCGAAAGATTCGGAGAAC 1519

DB 354 TTCTAGGAAATGGATGTTCCAAACTCTTCCACCTTACCCGAAAGATTCGGAGAAC 295

QY 1520 ATCTTCTAGCTATACGAAGTGGCTTGACAAAGAGTCAGGTATCAAACTGGTTTATAAATG 1579

DB 294 ATCTTCTAGCTATACGAAGTGGCTTGACAAAGAGTCAGGTATCAAACTGGTTTATAAATG 235

QY 1580 CGCGGTTAGGCTATGGAAGCCGATGATAGAGAGATGATGCGGAATGAACAAGAGA 1639

DB 234 CGCGGTTAGGCTATGGAAGCCGATGATAGAGAGATGATGCGGAATGAACAAGAGA 175

QY 1640 AGCTCAATAACAGTCACATTCATCAACCAACGACCACTTCCGAATGCCAAATCTGTGA 1699

DB 174 AGCTCAATAACAGTCACATTCATCAACCAACGACCACTTCCGAATGCCAAATCTGTGA 115

QY 1700 TGATGAGCCAAAGATGATAGAAATGAACAACAATGTTGTTTACCACTTGTGATAAT 1759

DB 114 TGATGAGCCAAAGATGATAGAAATGAACAACAATGTTGTTTACCACTTGTGATAAT 55

QY 1760 AGGCAATGCTACTACTATGATGCGCAACCACTTAACCATGTACCATCTATCAT 1813

DB 54 AGGCAATGCTACTACTATGATGCGCAACCACTTAACCATGTACCATCTATCAT 1

RESULT 3
AV808362/c
LOCUS
DEFINITION
mRNA sequence.
ACCESSION
AV808362
AV808362
RAFL9 Arabidopsis thaliana cDNA clone RAFL09-54-G14 3',
linear mRNA 423 bp
EST 29-MAR-2002

Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.

Carpio, T., Policky, J., Suzuki, G., Argentin, C., Shah, S.,
Nobrega, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and
Hanson, D.
Arabidopsis thaliana Gene Expression MicroArray
Unpublished (1999)

FEATURES source

Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.

ORIGIN

Query Match 26.9%; Score 508; DB 9; Length 511;

Best Local Similarity 99.4%; Pred. No. 2.8e-86;

Matches 508; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1210 CATGTCGAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCCTCCAAACCGTTTCC 1269

DB 511 CATGTCGAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCCTCCAAACCGTTTCC 452

QY 1270 TTCTTATACAGAACTGAGAGAGAGATCTGCAAGAGATAATCTCTATGGATCTGTA 1329

DB 451 TTCTTATACAGAACTGAGAGAGAGATCTGCAAGAGATAATCTCTATGGATCTGTA 392

QY 1330 TTGGAGAGGCAAGACCAAGACTCAAGAACTCTATGTTCCACGAGATTGCTTCTT 1389

DB 391 TTGGAGAGGCAAGACCAAGACTCAAGAACTCTATGTTCCACGAGATTGCTTCTT 332

QY 1390 CAGCAGCTGAACGAAAGAACCATCAGATTGGAGACCTCAACGAGGTTTCCCTGAGAAA 1449

DB 331 CAGCAGCTGAACGAAAGAACCATCAGATTGGAGACCTCAACGAGGTTTCCCTGAGAAA 272

QY 1450 TCTGTTTGGTCTACGAAATGGATGTTCCAAACTCTTCCACCTTACCCGAAAGAT 1509

DB 271 TCTGTTTGGTCTACGAAATGGATGTTCCAAACTCTTCCACCTTACCCGAAAGAT 212

QY 1510 TCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAAGATCAGGTATCAAACTGG 1569

DB 211 TCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAAGATCAGGTATCAAACTGG 152

QY 1570 TTTATAATGCGCGGTTAGCTATGGAAGCCGATGATAGAGAGATGATGCGGAATG 1629

DB 151 TTTATAATGCGCGGTTAGCTATGGAAGCCGATGATAGAGAGATGATGCGGAATG 92

QY 1630 AACAGAGGAAGCTCAATAACAGTCACATTCATCAACCAACGACCACTTCCGAATGCCA 1689

DB 91 AACAGAGGAAGCTCAATAACAGTCACATTCATCAACCAACGACCACTTCCGAATGCCA 32

QY 1690 AAATCTGTTATGATGAGCCAAAGCAATGCATA 1720

DB 31 AAATCTGTTATGATGAGCCAAAGCAATGCATA 1

RESULT 2 LOCUS

DEFINITION
thaliana cDNA clone 701557547, rosette-3 Arabidopsis
mRNA sequence.

ACCESSION
AI999808

VERSION
AI999808.1 GI:5946713

KEYWORDS
EST.

SOURCE
Arabidopsis thaliana (thale cress)

ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS
Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,
Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P.,
Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D.,
Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C.,

```

VERSION AV808362.1 GI:19842347
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 423)
AUTHORS Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.,
and Shinozaki, K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4339
Fax: 81-298-36-9060
Email: msek@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified lambda FLC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified plasmid vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES             Location/Qualifiers
     1..423
         /organism="Arabidopsis thaliana"
         /mol_type="mRNA"
         /db_xref="taxon:3702"
         /clone="RAF109-54-G14"
         /dev_stage="plants at various developmental stages from
         germination to mature seeds"
         /lab_host="DH10B"
         /clone_lib="RAP19"
         /note="Site 1: BamHI; Site 2: SalI; subjected to
         dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
         hr) treatments"

ORIGIN
Query Match      21.5%; Score 406.2; DB 9; Length 423;
Best Local Similarity 99.1%; Pred. No. 5.9e-67;
Matches 419; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1434 AGGTTTCCTCGAGAAATCTGTTTCGGTTCTACGGAATTGGATTTCCAAAACCTTCCTTCA 1493
DB 423 AGGTTTCCTCGAGAAATCTGTTTCGGTTCTACGGAATTGGATTTCCAAAACCTTCCTTCA 364
QY 1494 CCCTTACCCGAAGATTCGGAGAACATCTTCTAGCTATACGAAGTGGCTTGACAGAG 1553
DB 363 CCCTTACCCGAAGATTCGGAGAACATCTTCTAGCTATACGAAGTGGCTTGACAGAG 304
QY 1554 TCAGGTATCAAACTGGTTTAAATCGCGGGTAGGCTATGGAAGCCGATGATAGAAGA 1613
DB 303 TCAGGTATCAAACTGGTTTAAATCGCGGGTAGGCTATGGAAGCCGATGATAGAAGA 244
QY 1614 GATGATTCGGAATGAAACAGAGAGAGCTCAATACAGTCACATTCACCCCAACGACC 1673
DB 243 GATGATTCGGAATGAAACAGAGAGAGCTCAATACAGTCACATTCACCCCAACGACC 184
QY 1674 AACTCTTCGAATGCAAAATCTGTTATGATGAGCAGCAATGCATTAATAGACACAA 1733
DB 183 AACTCTTCGAATGCAAAATCTGTTATGATGAGCAGCAATGCATTAATAGACACAA 124
QY 1734 TTGTTTACCAACTTTGTGATTAATAGGCAATGTCTCTATGATGTCGCAAAACCTTA 1793
DB 123 TTGTTTACCAACTTTGTGATTAATAGGCAATGTCTCTATGATGTCGCAAAACCTTA 64
QY 1794 AACCATGACGACTATCAATGATGTTATTA-TTGTATATACAACTCCTTTATCTTTG 1852

```

```

DB 63 AACCATGTACGACTATCATTACGTATGTTATAAGTTGTATATACAACTCCTTTATCTTTG 4
QY 1853 ACT 1855
DB 3 ACT 1
RESULT 4
LOCUS CB261894
DEFINITION 83-E8862-008-010-K04-pbl2 MP12 ADIS-008 Arabidopsis thaliana cDNA
VERSION CB261894
KEYWORDS CB261894.1 GI:32886667
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 390)
AUTHORS Schmid, K.J., Soerensen, T.R., Stracke, R., Torjek, O., Altmann, T.,
Mitchell-Olds, T. and Weishaar, B.
TITLE Large-scale identification and analysis of genome-wide
single-nucleotide polymorphisms for mapping in Arabidopsis thaliana
JOURNAL Genome Res. 13 (6), 1250-1257 (2003)
MEDLINE 22883290
PUBMED 12799357
COMMENT Contact: Weishaar B
ADIS DNA core facility at MP12
Max-Planck-Institute for Plant Breeding Research
Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
Email: weishaar@mpiz-koeln.mpg.de
Insert Length: 390 Std Error: 0.00
Plate: 10 row: K column: 04
Seq primer: pB12; GGTGGCGCGCTCTAG.

FEATURES             Location/Qualifiers
     1..390
         /organism="Arabidopsis thaliana"
         /mol_type="mRNA"
         /cultivar="C24"
         /db_xref="GABI:601141"
         /db_xref="taxon:3702"
         /clone="MP12p76K0410Q"
         /tissue_type="seedling"
         /dev_stage="few days old seedlings"
         /lab_host="E. coli XLI-Blue MRF"
         /clone_lib="MP12-ADIS-008"
         /note="Vector: pBluescript SK (-); Site 1: EcoRI; Site 2:
         XhoI; cDNA library from Arabidopsis thaliana, accession
         C24; seedling: Lambda ZAPII phage library was made at the
         Max-Planck-Institute of Molecular Plant Physiology, Golem,
         Germany and mass-excised at the Max-Planck-Institute for
         Plant Breeding Research, Cologne, Germany; cloning sites
         EcoRI-XhoI; Note: Sequencing granted in the context of the
         'Establishment of high-efficiency SNP-based mapping tools
         and development of methods for genome-wide mutation
         detection', PI: Bernd Weishaar Sequence submission managed
         by RZPD/GABI-Primary database: http://gabi.rzpd.de This
         clone is available from RZPD; contact RZPD (clone@rzpd.de)
         for further information."

ORIGIN
Query Match      20.6%; Score 388.4; DB 14; Length 390;
Best Local Similarity 99.7%; Pred. No. 1.4e-63;
Matches 389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1436 GTTTCCTCGAGAAATCTGTTTCGGTTCTACGGAATGGATTTCCAAAACCTTCCTTCA 1495
DB 1 GTTTCCTCGAGAAATCTGTTTCGGTTCTACGGAATGGATTTCCAAAACCTTCCTTCA 60

```

1496 CTTACCCGAAAGATTCCGAGAAACATCTTCTAGCTATACGAAAGTGGCTTGACAAAGATC 1555
 Db CTTACCCGAAAGATTCCGAGAAACATCTTCTAGCTATACGAAAGTGGCTTGACAAAGATC 120
 QY 1556 AGGTATCAACTGCTTTTAAATGCGCGGTTAGGCTATGGAAGCCGATGATAGAGAGA 1615
 Db AGGTATCAACTGCTTTTAAATGCGCGGTTAGGCTATGGAAGCCGATGATAGAGAGA 180
 QY 1616 TGTATGCGGAAATGAACAGAGAGAGCTCAATAACAGTACACATTCACCCCAACGACCAA 1675
 Db TGTATGCGGAAATGAACAGAGAGAGCTCAATAACAGTACACATTCACCCCAACGACCAA 240
 QY 1676 CTCCTCGAATGCAAAATCTGTATGATGAGCCAGCAATGCAATTAATAAGCAACAATT 1735
 Db CTCCTCGAATGCAAAATCTGTATGATGAGCCAGCAATGCAATTAATAAGCAACAATT 300
 QY 1736 GTGTTTACCAACTTGTGATAATTAGGCAATTGCTACTCTATGATTGCCCAAAACCTAAA 1795
 Db GTGTTTACCAACTTGTGATAATTAGGCAATTGCTACTCTATGATTGCCCAAAACCTAAA 360
 QY 1796 CCATGACGACTATCATTTACGTTATTTATA 1825
 Db CCATGACGACTATCATTTACGTTATTTATA 390

RESULT 5

CA917399
 LOCUS EST641546 GPOD Medicago truncatula 773 bp mRNA linear EST 27-DEC-2002
 DEFINITION sequence.

CA917399

CA917399.1 GI:27404329

VERSION

KEYWORDS

SOURCE

ORGANISM

Medicago truncatula (barrel medic)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

1 (bases 1 to 773)

Grusak, M.A., Samac, D.A., Town, C.D., Van Aken, S., Utterback, T.,

Cheung, F. and Fraser, C.M.

More ESTs from developing reproductive tissues of Medicago

truncatula

Unpublished (2002)

Contact: Grusak, M.A.

USDA/ARS Children's Nutrition Research Center

Baylor College of Medicine

1100 Bates Street, Houston, TX 77030-2600, USA

Tel: 713 798 7044

Fax: 713 798 7078

Email: mgrusak@bcm.tmc.edu

TIGR sequence name: MTO3V25TK

More information is available at: www.medicago.org

Seq primer: SKmod (CTA GAA CTA gtc gat CC).

Location/Qualifiers

1. .773

/organism="Medicago truncatula"

/mol_type="mRNA"

/culivar="A17"

/db_xref="taxon:3880"

/clone="GPOD-3e51"

/dev_stage="Immature pod walls"

/days_after_pollination="15 to 30"

/clone_lib="GPOD"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:

XhoI; Immature pods, ranging in age from 15 to 30 days

after pollination, were collected from greenhouse-grown

plants. At harvest, seeds were removed from pods and

isolated pod walls were collected and immediately frozen

in liquid nitrogen. Pod walls were pooled for mRNA

extraction. cDNA was prepared from polyA+ enriched RNA.

The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XL04R cells."

ORIGIN

Query Match 14.7%; Score 277.6; DB 14; Length 773;
 Best Local Similarity 69.9%; Pred. No. 1.2e-42;
 Matches 411; Conservative 0; Mismatches 159; Indels 18; Gaps 2;
 QY 1068 TAACCTGGAGCGGGTTTCGGATCTACATTTCAAGAGAGCAATTAGAGCAAGAAAAC 1127
 Db 6 TATGTCAGAGCTTATGCTGATTTCTTTCGCAAGGCGATGCTGTTGAATCAAGCAATC 65
 QY 1128 CCATCTCTTGGATCTTCTTCAAAATGGTGGATGATCGATATAGTATGCTGGTATGAGAT 1187
 Db 66 TCAACCTCTTGGATCTTCTTCAAAATGGTGGATGATCGATATAGTATGCTGGTATGAGAT 125
 QY 1188 TCATACGGTTATATACAGGTTCCATGCTGCAACCGAGTTAGATCCACAGTTACACCCG 1247
 Db 126 TCACACTGTGCTATCTGCAATTTTCATGCTGCTGATCGATCCCAAAATACACCGCA 185
 QY 1248 GTTTGCGCTCCAAACCGTTTCTTCTTATACAAAGACCTGAGAGAGAGATCTTCAAGAA 1307
 Db 186 CTTTGCTGTCAAACCGTTTCTTCTTATACAAAGACCTGAGAGAGAGATCTTCAAGAA 245
 QY 1308 GATAATCTCTATGGGATCTGTATGGAGAG-----AGGCAAGACAGAAC 1352
 Db 246 TATCTTTCCATGGGATCAAAATTTTAAAGTTCATGTCAGAGAGAGATTAAGAAATTATC 305
 QY 1353 TCAGAAACCTCTATGTTCCACCGAGATTGCCTTCTTACAGAGCTGAAACGAGAAACCA 1412
 Db 306 TGTGAAACCTCTATGTTCCACCGAGATTGCCTTCTTACAGAGCTGAAACGAGAAACCA 362
 QY 1413 TCAGATTGGAGACCTTCAACGAGGTTTGCTCTGAGAAATCTGTTTCGTTCTACGGAATTG 1472
 Db 363 TCAGTTATGAGGCCCCCAAGGGGCTTCCAGAAAGATCTGTTTCGTTCTACGTTGATG 422
 QY 1473 GATGTTCCAAACTCTCTTCAACCGGTTACCGGAGAGATTCGAGAAACATCTTCTAGCTAT 1532
 Db 423 GATGTTTCGAAATTTCTCCATCCGATTCCTTAAAGATGAGAGAGAGATTTACTTCAAT 482
 QY 1533 ACAGAGTGGCTTGACAAAGAGTCAAGTATCAAACTGGTTTATAAATGCGCGGTTAGGCT 1592
 Db 483 AAAAAAGTGGTTGACTAGAGTCAAGTATCAAACTGGTTTATAAATGCGCGGTTAGGCT 542
 QY 1593 ATGGAAGCGGATGATAGAGAGATGATTCGGAATGAAACAGAGAGAA 1640
 Db 543 ATGGAACCACTGATTGAGGAAATGATGCTGAAATGAATAGAGAGAA 590

RESULT 6

BH476676

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Brassica oleracea

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 535)

Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other_GSSs: BGN8787T

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

BH476676
 BGN8787T BGN Brassica oleracea genomic clone BGN8787, genomic
 survey sequence.
 BH476676
 BH476676.1 GI:17684787
 GSS
 Brassica oleracea

Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 535)
 Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)

Other_GSSs: BGN8787T
 Contact: Chris Town

TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.

Email: akozik@atgc.org [michelm@vegmil.ucdavis.edu]
 belongs to [contig_QG_CA_Contig4128](#), see <http://cgpdb.ucdavis.edu/> for details.

```

source
1. .775
/organism="Lactuca sativa"

```

```

/ncr_type="nausa
/cultivar="L.serriola"
/db xref="taxon:4236"

```

```
/clone="QGF19G17"
```

```

/clone_note="QG_EFGHU lett
/clone_lib="QG_EFGHU lett
/note="vector: pBRCDNASfla
from 10 different sources

```

separate cDNAs were generated by incorporating unique 5' and 3' ends into the source of RNA. cDNAs were directionally cloned into

transformations made with size bias. Details of each construction can be obtained from TAC tissue-flowers pre-fer-

ORIGIN

Query Match 13.9%; Score 262.2;
Best Local Similarity 67.9%; Pred. No. 9.8
Matches 382; Conservative 0; Mismatches

QY
1082 GTTTCGGATCTACATTTCAAAGGAGCATTAG
| | | | |
Db
60 GATTGTTCATGCTTTCAAAGGAGCAAGTGTCGCG

1142 TTCTTCAAATGGTGGATGATCGATATAGTCATT
 120 TACTAGAAATGGTGGATGAGCGGTATAATCAGT

QY 1202 CAGCGTTCATGCTGCAACCGAGTTAGATCCAC
Db 180 CGGCATTTTCATGCTGTGAACAGAAATGAAATCCGCG

D**b** 240 CGATTACGTTTTTCTACAAAGTTTAAGAGAGA

1322 GATCTGTATT---GGAGAGAGGCAAGACAAGATQY

D**b** 300 GGGCGGATTTTAGCACCATGGATCAAGGAGAAG

Qy 1379 ATTGCCTTCTTCAGCAGCTGAAACGAAAGAACCC

D_b 360 AATGGGCTCTTCAGCAGCTAAGAAAGAAAGATC

QY 1439 TGCCTGAGAAATCTGTTTCGGTCTACGGAATT

Db
420 TTCCGGAAAAATCTGTTTCTGTGTTACGTGCAT

QY 1499 ACCCGAAGATTCCGGAGAAACATCTTCTAGCTA

Db
480 ATCCAAAGGATGCGGAGAAACAATTATTGGCTG

Qy 1559 TATCAAACTGGTTTATAATGCGGGGTTAGGC

Db 540 TATCGAATTGGTTATATAATCGAGGGTGAGAC

QY 1619 ATGCGAAATGAACAGAGGAG 1641
 Db 600 ACTTAGAGATGAATAGGAGG 622

RESULT 8
 BZ033022 719 bp DNA linear GSS 09-OCT-2002
 LOCUS oeh45h05.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
 DEFINITION sequence.

ACCESSION BZ033022
 VERSION BZ033022.1 GI:23607215
 KEYWORDS GSS.

SOURCE
 ORGANISM
 Brassica oleracea
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE
 1 (bases 1 to 719)
 Delaunay, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,
 Nash, W., Rabinowicz, P.D. and Wilson, R.K.
 Whole genome shotgun reads from Brassica oleracea
 Unpublished (2002)

TITLE
 JOURNAL
 COMMENT
 Contact: Richard K. Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@watson.wustl.edu
 Plate: oeh45 row: h column: 05
 Seq primer: -21UpOT forward
 Class: shotgun
 High quality sequence start: 16
 High quality sequence stop: 551.
 Location/Qualifiers

FEATURES
 source
 1..719
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /db_xref="taxon:3712"
 /clone_lib="B.oleracea002"
 /note="Vector: pOW13; Whole genome shotgun library from
 flowering buds. DNA was purified from a crude nuclear
 prep using Brassica oleracea T01000DH3 buds provided by
 Thomas Osborn at the University of Wisconsin. Genomic
 DNA was provided by Pablo Rabinowicz (CSHL) and the
 shotgun library prepared at Washington University Genome
 Sequencing Center."

ORIGIN
 Query Match 13.6%; Score 256.4; DB 28; Length 719;
 Best Local Similarity 66.3%; Pred. No. 1.3e-38;
 Matches 456; Conservative 0; Mismatches 156; Indels 76; Gaps 3;

QY 789 TTCGATGATCTCGAGATGAAGTCCTTTGTCAGCTACAGATTAGCCTCAGAGCAAGC 848
 Db 31 TTCTGATGAGTCTCAGAGATAAGTCCTTATGCTCTACTAAATGACCTCAGAGCAAGC 90

QY 849 TTCTTGCAGCAAGACATTTCTAATAAGCTTCTTACTCAAGTT---TCTCTCAACT 905
 Db 91 TTGACCTCAGAGCAAGCTTTCTCAGAGCAAGACATTATATACGTCCTCTCATGT 150

QY 906 TATATTGGCTCAAAATACCTTCATCTGTGTCAGAAATATCTATCTATTTTCGGCGGATA 965
 Db 151 TATATTGGCTCAAAAGTATCTTCACTCTGTTCAAGAAATACTATCTCAATTCGTACATA 210

QY 966 CTCGGT-----CGATTATTCATCTCGAGCAAGCGAGTC 998
 Db 211 CTCCTCCACGGCTTAGAGATAATCTCTCAATGCTATTTTTCATCTCAAGAAATCTGATC 270

QY 999 AGGAGCTGCTAGTTCAGGCTTTTACTTCACGTTTGGAGAAATAACTGAGTTTCTTGATGG 1058
 Db 271 AAGGGCTGCTGGTTTCTGAGCTTTTACTTCAGTTTACGTTTAGAGAAATCAACGGAGTTTCTTGAAG 330

QY 1059 TCAATCTAATA-----CT 1072

Db 331 AGGTTTGTATCAAAAGACGAGCATTTAGAACAAAGAAACCAGCTCTTGGATCTTCTTCA 390
 QY 1073 CGGAGCGGGTTTCGGATCTACATTTCAAGAGGAGCATTAGAACAAAGAAACCCTC 1132
 Db 391 AATGGTATATATTCACACATATTTATATACATAAGATGGAGAAATGATTTTGAAGTTT 450
 QY 1133 TCTTGATCTTCTTCAATGGTGATGATCGATATAGTCAATTCGGTAGATGAGATTCATA 1192
 Db 451 GAGTTGGATTTGCTCTCAGTGGATGATAATATAGTCAATTCGGTTGAGAGATTCATA 510
 QY 1193 CGTTATATCAGGTTCCATGTCGCAACCGAGTTAGATCCACAGTTTACACCCGGTTTG 1252
 Db 511 CGGTTGTATCAGCCTTCCATGCAGCAACCGAGTTAGACCCGCGAGCTACACCCGGTTTG 570
 QY 1253 CCCTCAAAACCGTTTCTTCTTATACAAGAACTGAGAGAGAGAAATCTTCAAGAGATAA 1312
 Db 571 CATCCAAACCATCTCGTTCTTATACAGAACTGAGGAGAGATTCAGCAAAAGATAC 630
 QY 1313 TCTTATGGATCTGTATTGGAGAGAGGCAAGAACTCAAGAACTCTATGTTCC 1372
 Db 631 TTATGATGGATCAGTTTATAGAGAGAGGCAAGAGAGTCTCAAGAAATTCATCATCC 690
 QY 1373 ACCAGATTGCCTTCTTCAGCAGCTGAA 1400
 Db 691 ACCAGATTGCCTTCTTCAGCAGATGAA 718

RESULT 9
 AV800635/c 270 bp mRNA linear EST 29-VAR-2002
 LOCUS AV800635 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-24-021 3,
 DEFINITION mRNA sequence.

ACCESSION AV800635
 VERSION AV800635.1 GI:19834620
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
 1 (bases 1 to 270)
 Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
 Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
 Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
 and Shinozaki, K.
 Large scale analysis of Arabidopsis full-length cDNA (2002b)
 Unpublished (2002)

TITLE
 JOURNAL
 COMMENT
 Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: msekietc@riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially
 as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
 and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
 al., submitted for publication) digested with BamHI and SalI. This
 clone is in a modified Bluescript vector. Please visit our web
 site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
 details.

FEATURES
 Location/Qualifiers
 1..270
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL09-24-021"
 /dev_stage="plants at various developmental stages from
 germination to mature seeds"
 /lab_host="DH10B"
 /clone_lib="RAFL9"
 /note="Site 1: BamHI; Site 2: SalI; subjected to
 dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24

hr) treatments"

ORIGIN

Query Match 13.3%; Score 251.6; DB 9; Length 270;
 Best Local Similarity 98.1%; Pred. No. 1.4e-37;
 Matches 265; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1587 TAGGCTATGAAGCCGATAGAGAGATGATGCGGAAATGAACAAGAGGAAGCTCAA 1646
 |||||
 DB 270 TAGGCTATGAAGCCGATAGAGAGATGATGCGGAAATGAACAAGAGGAAGCTCAA 211

QY 1647 TAACAGTCACATTCACCCACGACCACTCTTCGAATGCCAAATCTGTGTATGATGAG 1706
 |||||
 DB 210 TAACAGTCACATTCACCCACGACCACTCTTCGAATGCCAAATCTGTGTATGATGAG 151

QY 1707 CCAAGCAATGCATAAATAAGACAACAATTTGTTTACCAACTTTTGGATAA--TTAGGCAA 1765
 |||||
 DB 150 CCAAGCAATGCATAAATAAGACAACAATTTGTTTACCAACTTTTGGATAA--TTAGGCAA 91

QY 1766 TTGCTACTATGATTCGCCAAACCTTAACCACTATGACACTATCATTCAGTATGTTATA 1825
 |||||
 DB 90 TTGCTACTATGATTCGCCAAACCTTAACCACTATGACACTATCATTCAGTATGTTATA 31

QY 1826 ATTGTATATACAACTCTTTTATCTTTGACT 1955
 |||||
 DB 30 ATTGTATATACAACTCTTTTATCTTTGACT 1

RESULT 10

BU011811 676 bp mRNA linear EST 22-AUG-2002

LOCUS QGJ17D24.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone

DEFINITION QGJ17D24, mRNA sequence.

ACCESSION BU011811

VERSION BU011811.1 GI:22446206

KEYWORDS EST.

SOURCE Lactuca sativa

ORGANISM Lactuca sativa

REFERENCE

Ekaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
 Cichorieae; Lactuca.

1 (bases 1 to 676)

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
 Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
 Lai, Z., Church, S., Jackson, L. and Bradford, K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
<http://comgenomics.ucdavis.edu/>
 Unpublished (2002)

CONTACT: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.edu [michelmore@vegmil.ucdavis.edu]
 belongs to contig QG_CA_Config128, see <http://cgdb.ucdavis.edu/>
 for details.

Plate: QGJ17 row: D column: 24.

FEATURES

source

1..676
 Location/Qualifiers

1..676
 /organism="Lactuca sativa"
 /mol_type="mRNA"
 /culti_var="L.serriola"
 /db_xref="taxon:4236"
 /clone="QGJ17D24"
 /lab_host="E.coli"

NOTE: QG_EFGHJ lettuce serriola
 /notes="Vector: pBRCDNA5f1AB; The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each

source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at <http://cgdb.ucdavis.edu/>
 TAG_TISSUE=flowers pre-fertilized
 TAG_LIB=QG_EFGHJ lettuce serriola
 TAG_SEQ=GCTTGACGGG"

ORIGIN

Query Match 12.9%; Score 243; DB 13; Length 676;
 Best Local Similarity 69.9%; Pred. No. 4.5e-36;
 Matches 343; Conservative 0; Mismatches 145; Indels 3; Gaps 1;

QY 1154 TGGATGATCGATATAGTCATTGCGTAGATGAGATTACGTTATATACGCGTTCATG 1213
 |||||
 DB 1 TGGATGATCGATATAGTCATTGCGTAGATGAGATTACGTTATATACGCGTTCATG 60

QY 1214 CTGCAACCGAGTTAGATCCACAGTTACACCCCGTTTGCCCTCAACCGCTTCTTCT 1273
 |||||
 DB 61 CTGTAACAGAAATTGAATCCGCAAGTTTCATGCTGTTTGTCTCTTCATACGATTACGTTT 120

QY 1274 TATCAAGAACTCTGAGAGAGAGAACTCTGCAAGAGATAATCTCTATGGATCTGTATT-- 1331
 |||||
 DB 121 TCTACAAAAGTTTAAAGAGAGAGATAAGTAATTATATACCTTCAATGGGGCGGATTTA 180

QY 1332 -GGAGAGAGGCAAGACAAGACTCAAGAAACCTCTATGTTCCACAGCAATTCCTTCT 1390
 |||||
 DB 181 GCACCATGATCAAGGAGAGAGATGAGTGTCTGTTTGTACCAAGCAATGGGCTCTTC 240

QY 1391 AGCAGCTGAAACGAAAGAACCATCAGATTTGGAGACCTCACGAGGTTTGCTCGAGAAT 1450
 |||||
 DB 241 AGCAGCTGAAGAAAGAAAGATCATCAGCTATGGAGGCGCAAGAGGGCTTCGGAAAT 300

QY 1451 CTGTTTCGGTTCTACGGAATTGGATGTTTCCAAAACCTTCTTACCCCTTACCCGAAAGATT 1510
 |||||
 DB 301 CTGTTTCTGTGTACGTGCATGGATGTTTCAGAACTTCTTCCACCGTATCCAAAGGATG 360

QY 1511 CGGAAACATCTTCTAGCTATAGAAAGTGCCTTGACAGAGCTCAGGTATCAACATGCT 1570
 |||||
 DB 361 CGGAAACATCTTCTAGCTATAGAAAGTGCCTTGAAAGCCATGATGAGAAATGTACTTAGATGA 420

QY 1571 TTATAAATGCCGGGTTAGGCTATGGAAGCCGATGATAGAGAGATGTATCGGAAATGA 1630
 |||||
 DB 421 TTATAAATGCCGGGTTAGGCTATGGAAGCCGATGATGAGAAATGTACTTAGATGA 480

QY 1631 ACAAGAGGAAG 1641
 |||||
 DB 481 ATAGAGGAGG 491

RESULT 11

BU031456/c

LOCUS BZ031456.1 B.oleracea002 Brassica oleracea genomic, genomic survey

DEFINITION sequence.

ACCESSION BZ031456

VERSION BZ031456.1 GI:23604956

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

REFERENCE

1 (bases 1 to 709)
 Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,
 Nash, W., Rabinowicz, P.D. and Wilson, R.K.
 Whole genome shotgun reads from Brassica oleracea
 Unpublished (2002)
 Contact: Richard K. Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@watson.wustl.edu

```

Plate: cef98 row: b column: 09
Seq primer: -28RpPOT reverse
Class: Shotgun
High quality sequence start: 16
High quality sequence stop: 551.
Location/Qualifiers
1..709
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/notes="vector: pOTW3; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablic Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

```

Query Match	11.8%;	Score 222.4;	DB 28;	Length 709;
Best Local Similarity	69.0%;	Pred. No. 3.5e-32;		
Matches 486;	Conservative	0;	Mismatches 86;	Indels 132;
Gaps	8			

	496	AACAAATACTTATCATCTTTCCACGGGTCTATCGATTGTAAGATAACCGCAATGTTGAG	555
QY			
	708	AAAGAAGTACTTATCATCTACTTCTAAGATCTACAATGTTTCATGAACCAGC---	652
Db			
	556	TTCATGGGTCTCCCTCCCTCATCTCCCTCCACTTCATCTCTTTGGATCATTTTAAGACACTAT	615
QY			
	651	GCCATGGGTCC-----TCTCTCTTCACTTCTATGATCATCTTAGACCATAT	601
Db			
	616	GATGATCTCTCAAACAACATGTGGGGTTTTGAAGCAAAATAGTGAGTTTCAGGCATTTTC-	674
QY			
	600	GATGATCTCCTCAAAAACAACATGTGAGTCTTGGTGAAAAATAGTGAGTTTCATGTCATATCCA	541
Db			
	675	-----AGGTGTAGTTGGTCCAAGTGAAACCAATGATGTCACATTCGGTGAAGAAGAT	726
QY			
	540	TTCATAGAAGGGTGTGTTGGTCTTAGTGAAACCAATAATCTCACATTCGGTGAAGAA--	484
Db			
	727	TTCGCCGTTTCTAAATTCGAATAAAAGAAACAATGAGCTTTCAITGAGTCTTGCATCAGAT	786
QY			
	483	-----TTCCCAAGAAACGAGTTCTCGTTAAGTCTTGGCCACTGAT	445
Db			
	787	GTTTCTTGATGATGCTCGGAGATTAAGTCTTTGTGC-----AGCTACAGATTAAGCC	837
QY			
	444	GTTTCTTGATGATGCTCGGAGATTAAGTCTTTGTGCTGCTACTAAATCTACTAGAAATACC	385
Db			
	838	TCAGAGCAAGCTTCTTGACAGCAGCAAGAACACATTTCTAAATAACGTTGTTACTCAAGGTTTC	897
QY			
	384	TCAGAGCAAGCTTCTTCCAGCAGCAAGAACACATGTCAAAACG-----TC	343
Db			
	898	TCTCAACTTATATTGGCTCAAAATACCTTCACCTCTGTTCAAGAAATACTATCTCATTTTC	957
QY			
	342	TCTCAAGTCATATTTTGTCTCAAGTACCTTCACTCTGTTCAAGAAATACTATCTCATTTTC	283
Db			
	958	GCCGATACTCGCT-----CCATTATTCACTCTCGAGCA	990
QY			
	282	GCCATACTCCCTCAAGGCTTAGAGACTAATCCTCAGTGCTACTTTTCACTCGAGCA	223
Db			
	991	ACCGAGTCAGAGCTGTAGTTTCAGCCTTTACTTCACGTTTTCAGAAATATACTAGTTT	1051
QY			
	222	ACCGAGTCAGAGGTPACAACTCAAGCCTTTACTTCATGTTTATGAGAATCTTAAACGAGTTT	163
Db			
	1051	CTTGATGGTGATTCTTAATAACTCGAGCGGGTTTCGGATCTACATTTTCAAGGAGAGCA	1111
QY			
	162	CTTGATGCGGGTTC-----GCANAGACAAGCA	136
Db			
	1111	TTAGAAGCAAGAAACCCTACTCTTGATCTTCTTCAAAATGTT	1154
QY			
	135	TTAGAAGCAAGAAACCCTACTCTTGATCTTCTTCAAAATGTT	92
Db			

RESULT 12	REFERENCE
BU837920	AUTHORS
LOCUS	TITLE
DEFINITION	JOURNAL
ACCESSION	COMMENT
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

FEATURES	SOURCE
1. Age	1. Age
2. Gender	2. Gender
3. Marital Status	3. Marital Status
4. Education	4. Education
5. Income	5. Income
6. Occupation	6. Occupation
7. Religion	7. Religion
8. Political Affiliation	8. Political Affiliation
9. Health Status	9. Health Status
10. Travel History	10. Travel History
11. Employment Status	11. Employment Status
12. Family Size	12. Family Size
13. Home Ownership	13. Home Ownership
14. Vehicle Ownership	14. Vehicle Ownership
15. Insurance Status	15. Insurance Status
16. Charitable Donations	16. Charitable Donations
17. Volunteer Work	17. Volunteer Work
18. Substance Use	18. Substance Use
19. Mental Health	19. Mental Health
20. Physical Activity	20. Physical Activity
21. Dietary Habits	21. Dietary Habits
22. Smoking Status	22. Smoking Status
23. Alcohol Consumption	23. Alcohol Consumption
24. Stress Levels	24. Stress Levels
25. Social Media Usage	25. Social Media Usage
26. Travel Frequency	26. Travel Frequency
27. Home Security	27. Home Security
28. Vehicle Maintenance	28. Vehicle Maintenance
29. Insurance Premiums	29. Insurance Premiums
30. Charitable Giving	30. Charitable Giving
31. Volunteer Hours	31. Volunteer Hours
32. Substance Abuse	32. Substance Abuse
33. Mental Health Services	33. Mental Health Services
34. Physical Activity Levels	34. Physical Activity Levels
35. Dietary Restrictions	35. Dietary Restrictions
36. Smoking Cessation	36. Smoking Cessation
37. Alcohol Treatment	37. Alcohol Treatment
38. Stress Management	38. Stress Management
39. Social Media Engagement	39. Social Media Engagement
40. Travel Itinerary	40. Travel Itinerary
41. Home Safety Measures	41. Home Safety Measures
42. Vehicle Safety Features	42. Vehicle Safety Features
43. Insurance Claims	43. Insurance Claims
44. Charitable Impact	44. Charitable Impact
45. Volunteer Impact	45. Volunteer Impact
46. Substance Abuse Treatment	46. Substance Abuse Treatment
47. Mental Health Outcomes	47. Mental Health Outcomes
48. Physical Activity Benefits	48. Physical Activity Benefits
49. Dietary Health Effects	49. Dietary Health Effects
50. Smoking Cessation Success	50. Smoking Cessation Success
51. Alcohol Consumption Reduction	51. Alcohol Consumption Reduction
52. Stress Management Techniques	52. Stress Management Techniques
53. Social Media Influence	53. Social Media Influence
54. Travel Safety Incidents	54. Travel Safety Incidents
55. Home Security Incidents	55. Home Security Incidents
56. Vehicle Safety Incidents	56. Vehicle Safety Incidents
57. Insurance Premium Changes	57. Insurance Premium Changes
58. Charitable Giving Trends	58. Charitable Giving Trends
59. Volunteer Work Trends	59. Volunteer Work Trends
60. Substance Abuse Trends	60. Substance Abuse Trends
61. Mental Health Trends	61. Mental Health Trends
62. Physical Activity Trends	62. Physical Activity Trends
63. Dietary Trends	63. Dietary Trends
64. Smoking Trends	64. Smoking Trends
65. Alcohol Trends	65. Alcohol Trends
66. Stress Trends	66. Stress Trends
67. Social Media Trends	67. Social Media Trends
68. Travel Trends	68. Travel Trends
69. Home Security Trends	69. Home Security Trends
70. Vehicle Safety Trends	70. Vehicle Safety Trends
71. Insurance Trends	71. Insurance Trends
72. Charitable Trends	72. Charitable Trends
73. Volunteer Trends	73. Volunteer Trends
74. Substance Abuse Trends	74. Substance Abuse Trends
75. Mental Health Trends	75. Mental Health Trends
76. Physical Activity Trends	76. Physical Activity Trends
77. Dietary Trends	77. Dietary Trends
78. Smoking Trends	78. Smoking Trends
79. Alcohol Trends	79. Alcohol Trends
80. Stress Trends	80. Stress Trends
81. Social Media Trends	81. Social Media Trends
82. Travel Trends	82. Travel Trends
83. Home Security Trends	83. Home Security Trends
84. Vehicle Safety Trends	84. Vehicle Safety Trends
85. Insurance Trends	85. Insurance Trends
86. Charitable Trends	86. Charitable Trends
87. Volunteer Trends	87. Volunteer Trends
88. Substance Abuse Trends	88. Substance Abuse Trends
89. Mental Health Trends	89. Mental Health Trends
90. Physical Activity Trends	90. Physical Activity Trends
91. Dietary Trends	91. Dietary Trends
92. Smoking Trends	92. Smoking Trends
93. Alcohol Trends	93. Alcohol Trends
94. Stress Trends	94. Stress Trends
95. Social Media Trends	95. Social Media Trends
96. Travel Trends	96. Travel Trends
97. Home Security Trends	97. Home Security Trends
98. Vehicle Safety Trends	98. Vehicle Safety Trends
99. Insurance Trends	99. Insurance Trends
100. Charitable Trends	100. Charitable Trends

ORIGIN

Query Match	11.5%; Score 216.4; DB 13; Length 751;
Best Local Similarity	68.8%; Pred. No. 4.8e-31;
Matches 333; Conservative	0; Mismatches 136; Indels 15; Gaps 2;
Qy	1094 CATTTCAAAGGAGAGCATTAGAAGCAAGAAACCCATCTCTTGGATCTCTTCCAATGG 1153
Db	
Qy	271 CATGCAGAAACGGGCACCTGGAAGCAAGAGAAACAAATGCTAATCTACTCCAAGTGG 330
Db	
Qy	1154 TGGATGATCGATATAGTCTCATTTGGTATAGATGAGATTCATACGGTTATATCAGCGTTCATG 1213
Db	
Qy	331 TTGATGAACGATATAGCCAAATGCTTGGATGAATAACACAGATTATTTGAGCATTCATG 390
Db	
Qy	1214 CTCGAACCGAGTTAGATTCACAGATTACACACCGGTTTGGCCTCCAAACCGTTTCCTTCT 1273
Db	
Qy	391 CTGCAACTGAGTTGGACCCACAGATTACACACAGTTTTTCTCTTCCAACAATAATCTTTCT 450
Db	
Qy	1274 TATACAAGAACCTCGAGAGAGAGAAATCGCAAGAAGATAATCTCTATGGGATCTGTATTGG 1333
Db	
Qy	451 TGTACAACAGCCTTAGGGAGCGGATCAGAACCAATCTTAGCAATGGGAGCTCATTTGG 510
Db	
Qy	1334 AGAG-----AGGCAAGACACAAGACTCAAGAAACCTCTATGTTTCCACCAGATT 1381
Db	
Qy	511 ATAGTGGAGACACCATAGAGACAGAAGGATCTTTTGGAACTTCATACTTCAAAAAGCAAT 570
Db	
Qy	1382 GCCTTTCTCAGCAGCTGAACGAAAGAACCATCAGATTTGGAGACCTCAACGAGGTTTGC 1441
Db	
Qy	571 GGACTCTCAGCAGCTGAAGAAAAATGACCAAC---TGTGGAGACCCGAGAGAGGCTTAC 627
Db	
Qy	1442 CTGAGAAATCTGTTTCGGTTCTACGGAATTTGGATGTTTCCAAAACCTTCTTTCACCCCTTACC 1501
Db	
Qy	628 CAGAAAGATCTGTCTCAGTTCTCGGTGCGATGGATGTTTTCAGAACTTTCTTACCCGTACC 687
Db	
Qy	1502 CGAAAGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTTGACAAGAGTTCAGGTAT 1561
Db	
Qy	688 CTAAGATCAGAGAAGCATTTGCTTGGCGCAAAAAGTGGACTTAAACAAGAGCCAGGTAT 747
Db	
Qy	1562 CAAA 1565
Db	
Qy	748 CAAA 751
Db	

RESULT 13	BU024863	672 bp	mRNA	linear	EST 23-AUG-2002
LOCUS	QHFGH21.yg.ab1	QH_EFGHJ sunflower RHA280	Helianthus annuus	cDNA	
DEFINITION	clone QHF6H21, mRNA sequence.				
ACCESSION	BU024863				
KEYWORDS	BU024863.1	GI:22460383			
SOURCE	EST.				
ORGANISM	Helianthus annuus (common sunflower)				
REFERENCE	<p>Helianthus annuus Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asteraceae; Asteroideae; Heliantheae; Helianthus. 1 (bases 1 to 672)</p>				
AUTHORS	<p>Kozik, A., Michelmore, R. W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Elliott, P., Kolkman, J., Slabaugh, M. S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K. Letuce and Sunflower ESTs from the Compositae Genome Project http://comgenomics.ucdavis.edu/ Unpublished (2002) Contact: Alexander Kozik [R.W.Michelmore] Department of Vegetable Crops, R.W.Michelmore Lab University of California at Davis (UCD) Asmunsdon Hall, UCD, Davis, CA 95616, USA Tel: 1-(530)-742-1742 Fax: 1-(530)-752-9659 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu] belongs to contig QH_CA_Contig2963, see http://cgdb.ucdavis.edu/ for details. Plate: QHF6 row: H column: 21.</p>				
TITLE	<p>Location/Qualifiers 1. .672 /organism="Helianthus annuus" /mol_type="mRNA" /cultivar="RHA280" /db_xref="taxon:4232" /clone="QHF6H21" /lab_host="E.coli" /clone_lib="QH_EFGHJ sunflower RHA280" /notes="Vector: pBRCDNASFIAB; The library was constructed from 11 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/ TAG_TISSUE=hulls TAG_LIB=QH_EFGHJ sunflower RHA280 TAG_SEQ=GCTAGTCGGG"</p>				
JOURNAL					
COMMENT					
FEATURES					
source					
ORIGIN					
Query Match	7.8%; Score 147.4; DB 13; Length 672;				
Best Local Similarity	69.8%; Pred. No. 6e-18;				
Matches 199; Conservative	0; Mismatches 86; Indels 0; Gaps 0;				
QY	1368	GTTCACACAGATTCGCTTCTTCAGCAGCTGAACGAAAGAACCATCAGATTGGAGACC	1427		
Db	80	GCTCCGGCAACACGGCGCTTTACAGCAATTAGGAATGATCCACACAAATGCTTCGAGACC	139		
QY	1428	TCAACGAGGTTTGCTCAGAAATCTGTTTCGGTTCTACGGAATGGATGTTCCTCAAAACTT	1487		
Db	140	ACAAGAGGATGCGGAAACGCTCTGTTTCGGTTCTTCGTTGGCTGTTGAACACTT	199		
QY	1488	CTTTCACCCCTTACCAGAAAGATTGCGAGAAACATCTTCTCTAGCTATACGAAGTGCCTTAC	1547		
Db	200	TCTCCATCCCTTATCCAAAGATTGCGATAAACAATGCTGTCGCAAAACAAATCGGACTCAC	259		

QY 1488 CCTTACCCCTTACCGAAGATTCGGAGAAACATCTTTAGCTATACGAAGTGGCTTGC 1547
 Db 228 TCTCCATCCTTATCCAAAGATTTCGATAAAACATATGCTGGCAAAACAACCTGGACTCAC 287
 QY 1548 AAGAAGTCAGGTATCAAACTGCTTTTATAAATCGCGGGTTAGGCTATCGAAGCCGATGAT 1607
 Db 288 GAGAAGTCAGGTTCAAACTGCTTTTATAAATCGTATAGAGTTCGGCTATGGAAGCCCAATGGT 347
 QY 1608 AGAAGAGATGTATCGGAAATGAACAAGAGGAGTCAATAACAG 1652
 Db 348 AGAAGAGATGTACTTGGAGAATGCAAGGAACATGAACAGACAG 392

RESULT 15

AL946255
 LOCUS 328 bp DNA linear GSS 24-OCT-2002
 DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-295B12-015445,
 genomic survey sequence.
 ACCESSION AL946255
 VERSION AL946255.1 GI:24402877
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 REFERENCE 1 Strizhov,N., Li,Y., Rosso,M., Viehoveer,P., Dekker,K., Saedler,H.
 and Weisshaar,B.
 A pipeline for automated high-throughput generation of FSTs
 (flanking sequence tags) from Arabidopsis thaliana T-DNA
 transformed lines
 Unpublished
 JOURNAL
 REFERENCE 2 Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
 A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
 for flanking sequence tag based reverse genetics
 Unpublished
 JOURNAL
 REFERENCE 3 (bases 1 to 328)
 Li,Y., Strizhov,N., Rosso,M. and Weisshaar,B.
 Direct Submission
 Submitted (21-OCT-2002) Weisshaar B., Max-Planck-Institut fuer
 Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
 This sequence is recovered from the left border of the T-DNA. It
 indicates an insertion within the locus defined by clone F26P21.
 The sequences are generated at the MPI for Plant Breeding Research
 in the context of the GABI-Kat project. GABI-Kat is part of the
 German Plant Genomics program designated 'GABI'. Information on
 line availability can be found at:
 http://www.mpiz-koeln.mpg.de/GABI-Kat/.

FEATURES

source
 1..328
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
 /note="PCR was performed on DNA from Arabidopsis thaliana
 plants (T1) which were transformed with the T-DNA from
 vector PAC161. The lines contain one or more T-DNA
 insertions. The DNA fragment(s) resulting from the PCR
 were directly sequenced to determine the genomic sequence
 flanking the insertion. Sequences displaying significant
 similarity to the A. thaliana nuclear genome sequence were
 processed for submission. T-DNA derived sequences were
 removed"

ORIGIN

Query Match 7.7%; Score 146; DB 29; Length 328;
 Best Local Similarity 98.8%; Pred. No. 1.4e-17;
 Matches 168; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 119 CTGCGAATGAAAGTACACTGTTCTTTTCACGGAGAAAGAGATAAATAAGCATTATCTTC 178
 Db 25 CTGCGAATGAAAGT-CACGTGTTCTTTTCACGGAGAAAGAGATAAATAAGCATTATCTTC 83
 QY 179 TTCTTTCAGTTTAAACACACACATTTTGGAAATTTTGATGTAAAAATCTCTTTTGGACGTT 238
 Db 84 TTCTT-AGTTTTTAAACACACATTTTGGAAATTTTGATGTAAAAATCTCTTTTGGACGTT 142
 QY 239 GTGTTGTCTGAAATCTTCCAAAGGTTCTATCAGAAAGAAAGGATAAAG 288
 Db 143 GTGTTGTCTGAAATCTTCCAAAGGTTCTATCAGAAAGAAAGGATAAAG 192

Search completed: September 2, 2004, 09:48:35
 Job time : 4997 secs